

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
6 July 2000 (06.07.2000)

PCT

(10) International Publication Number
WO 00/39304 A2

(51) International Patent Classification⁷: C12N 15/49, A61K 48/00

(21) International Application Number: PCT/US99/31273

(22) International Filing Date:
30 December 1999 (30.12.1999)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/114,495 31 December 1998 (31.12.1998) US
60/152,195 1 September 1999 (01.09.1999) US

(71) Applicant: CHIRON CORPORATION [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US).

(72) Inventors: BARNETT, Susan; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). ZUR MEGEDE, Jan; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US).

(74) Agents: DOLLARD, Anne, S.; Chiron Corporation, Intellectual Property - R440, P.O. Box 8097, Emeryville, CA 94662-8097 et al. (US).

(81) Designated States (*national*): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— Without international search report and to be republished upon receipt of that report.

(48) Date of publication of this corrected version:
7 December 2000

(15) Information about Correction:
see PCT Gazette No. 49/2000 of 7 December 2000, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF

(57) Abstract: The present invention relates to polynucleotides encoding immunogenic HIV type C Gag- and/or Env-containing polypeptides. Uses of the polynucleotides in applications including DNA immunization, generation of packaging cell lines, and production of Gag- and/or Env-containing proteins are also described.

WO 00/39304 A2

POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C POLYPEPTIDES,
POLYPEPTIDES AND USES THEREOF

TECHNICAL FIELD

5 Polynucleotides encoding antigenic Type C HIV Gag-containing polypeptides and/or Env-containing polypeptides are described, as are uses of these polynucleotides and polypeptide products in immunogenic compositions.

BACKGROUND OF THE INVENTION

10 Acquired immune deficiency syndrome (AIDS) is recognized as one of the greatest health threats facing modern medicine. There is, as yet, no cure for this disease.

In 1983-1984, three groups independently identified the suspected etiological agent of AIDS. See, e.g., Barre-Sinoussi et al. (1983) Science 220:868-871; Montagnier et al., in Human T-Cell Leukemia Viruses (Gallo, Essex & Gross, eds., 1984); Vilmer et al. (1984) The Lancet 1:753; Popovic et al. (1984) Science 224:497-500; Levy et al. (1984) Science 225:840-842. These isolates were variously called lymphadenopathy-associated virus (LAV), human T-cell lymphotropic virus type III (HTLV-III), or AIDS-associated retrovirus (ARV). All of these isolates are strains of the same virus, and were later collectively named Human Immunodeficiency Virus (HIV). With the isolation of a related AIDS-causing virus, the strains originally called HIV are now termed HIV-1 and the related virus is called HIV-2 See, e.g., Guyader et al. (1987) Nature 326:662-669; Brun-Vezinet et al. (1986) Science 233:343-346; Clavel et al. (1986) Nature 324:691-695.

25 A great deal of information has been gathered about the HIV virus, however, to date an effective vaccine has not been identified. Several targets for vaccine development have been examined including the *env* and *Gag* gene products encoded by HIV. Gag gene products include, but are not limited to, Gag-polymerase and Gag-

protease. Env gene products include, but are not limited to, monomeric gp120 polypeptides, oligomeric gp140 polypeptides and gp160 polypeptides.

Haas, et al., (*Current Biology* 6(3):315-324, 1996) suggested that selective codon usage by HIV-1 appeared to account for a substantial fraction of the inefficiency of viral protein synthesis. Andre, et al., (*J. Virol.* 72(2):1497-1503, 1998) described an increased immune response elicited by DNA vaccination employing a synthetic gp120 sequence with optimized codon usage. Schneider, et al., (*J Virol.* 71(7):4892-4903, 1997) discuss inactivation of inhibitory (or instability) elements (INS) located within the coding sequences of the Gag and Gag-protease coding sequences.

The *Gag* proteins of HIV-1 are necessary for the assembly of virus-like particles. HIV-1 *Gag* proteins are involved in many stages of the life cycle of the virus including, assembly, virion maturation after particle release, and early post-entry steps in virus replication. The roles of HIV-1 *Gag* proteins are numerous and complex (Freed, E.O., *Virology* 251:1-15, 1998).

Wolf, et al., (PCT International Application, WO 96/30523, published 3 October 1996; European Patent Application, Publication No. 0 449 116 A1, published 2 October 1991) have described the use of altered pr55 *Gag* of HIV-1 to act as a non-infectious retroviral-like particulate carrier, in particular, for the presentation of immunologically important epitopes. Wang, et al., (*Virology* 200:524-534, 1994) describe a system to study assembly of HIV *Gag*- β -galactosidase fusion proteins into virions. They describe the construction of sequences encoding HIV *Gag*- β -galactosidase fusion proteins, the expression of such sequences in the presence of HIV *Gag* proteins, and assembly of these proteins into virus particles.

Shiver, et al., (PCT International Application, WO 98/34640, published 13 August 1998) described altering HIV-1 (CAM1) *Gag* coding sequences to produce synthetic DNA molecules encoding HIV *Gag* and modifications of HIV *Gag*. The codons of the synthetic molecules were codons preferred by a projected host cell.

Recently, use of HIV Env polypeptides in immunogenic compositions has been described. (see, U.S. Patent No. 5,846,546 to Hurwitz et al., issued December 8, 1998, describing immunogenic compositions comprising a mixture of at least four different recombinant virus that each express a different HIV env variant; and U.S.

Patent No. 5,840,313 to Vahlne et al., issued November 24, 1998, describing peptides which correspond to epitopes of the HIV-1 gp120 protein). In addition, U.S. Patent No. 5,876,731 to Sia et al, issued March 2, 1999 describes candidate vaccines against HIV comprising an amino acid sequence of a T-cell epitope of Gag linked directly to an amino acid sequence of a B-cell epitope of the V3 loop protein of an HIV-1 isolate containing the sequence GPGR. There remains a need for antigenic HIV polypeptides, particularly Type C isolates.

SUMMARY OF THE INVENTION

The present invention relates to improved expression of HIV Type C *Gag*-containing polypeptides and production of virus-like particles, as well as, *Env*-containing polypeptides.

One aspect of the present invention relates to expression cassettes and polynucleotides contained therein. In one embodiment, an expression cassette comprises a polynucleotide sequence encoding a polypeptide including an HIV *Gag*-containing polypeptide, wherein the polynucleotide sequence encoding the *Gag* polypeptide comprises a sequence having at least about 85%, preferably about 90%, more preferably about 95%, and most preferably about 98% sequence identity to the sequences taught in the present specification. The polynucleotide sequences encoding *Gag*-containing polypeptides include, but are not limited to, the following polynucleotides: nucleotides 844-903 of Figure 1 (a *Gag* major homology region) (SEQ ID NO:1); nucleotides 841-900 of Figure 2 (a *Gag* major homology region) (SEQ ID NO:2); the sequence presented as Figure 1 (SEQ ID NO:3); and the sequence presented as Figure 2 (SEQ ID NO:4). The polynucleotides encoding the *Gag*-containing polypeptides of the present invention may also include sequences encoding additional polypeptides. Such additional polynucleotides encoding polypeptides may include, for example, coding sequences for other HIV proteins, such as, polynucleotide sequences encoding an HIV *protease* polypeptide, and polynucleotide sequences encoding an HIV *polymerase* polypeptide. In one embodiment, the sequence encoding the HIV *polymerase* polypeptide can be modified by deletions of coding regions corresponding to reverse transcriptase and integrase. Such deletions in the polymerase polypeptide can also be made such that the

polynucleotide sequence preserves T-helper cell and CTL epitopes. Other antigens of interest may be inserted into the polymerase as well.

In another embodiment, an expression cassette comprises a polynucleotide sequence encoding a polypeptide including an HIV *Env*-containing polypeptide, wherein the polynucleotide sequence encoding the *Gag* polypeptide comprises a sequence having at least about 85%, preferably about 90%, more preferably about 95%, and most preferably about 98% sequence identity to the sequences taught in the present specification. The polynucleotide sequences encoding *Env*-containing polypeptides include, but are not limited to, the following polynucleotides:

10 nucleotides 1213-1353 of Figure 3 (SEQ ID NO:5) (an *Env* common region); nucleotides 82-1512 of Figure 3 (SEQ ID NO:6) (a gp120 polypeptide); nucleotides 82-2025 of Figure 3 (SEQ ID NO:7) (a gp140 polypeptide); nucleotides 82-2547 of Figure 3 (SEQ ID NO:8) (a gp160 polypeptide); nucleotides 1-2547 of Figure 3 (SEQ ID NO:9) (a gp160 polypeptide with signal sequence); nucleotides 1513-2547 of Figure 3 (SEQ ID NO:10) (a gp41 polypeptide); nucleotides 1210-1353 of Figure 4 (SEQ ID NO:11) (an *Env* common region); nucleotides 73-1509 of Figure 4 (SEQ ID NO:12) (a gp120 polypeptide); nucleotides 73-2022 of Figure 4 (SEQ ID NO:13) (a gp140 polypeptide); nucleotides 73-2565 of Figure 4 (SEQ ID NO:14) (a gp160 polypeptide); nucleotides 1-2565 of Figure 4 (SEQ ID NO:15) (a gp160 polypeptide with signal sequence); and nucleotides 1510-2565 of Figure 4 (SEQ ID NO:16) (a gp41 polypeptide).

15 20

The present invention further includes recombinant expression systems for use in selected host cells, wherein the recombinant expression systems employ the polynucleotides and expression cassettes of the present invention. In such systems, the polynucleotide sequences are operably linked to control elements compatible with expression in the selected host cell. Numerous expression control elements are known to those in the art, including, but not limited to, the following: transcription promoters, transcription enhancer elements, transcription termination signals, polyadenylation sequences, sequences for optimization of initiation of translation, and translation termination sequences. Exemplary transcription promoters include, but are not limited to those derived from CMV, CMV+intron A, SV40, RSV, HIV-Ltr, MMLV-ltr, and metallothionein.

25 30

In another aspect the invention includes cells comprising the expression cassettes of the present invention where the polynucleotide sequence (e.g., encoding an Env- and/or Gag-containing polypeptide) is operably linked to control elements compatible with expression in the selected cell. In one embodiment such cells are

5 mammalian cells. Exemplary mammalian cells include, but are not limited to, BHK, VERO, HT1080, 293, RD, COS-7, and CHO cells. Other cells, cell types, tissue types, etc., that may be useful in the practice of the present invention include, but are not limited to, those obtained from the following: insects (e.g., *Trichoplusia ni* (Tn5) and Sf9), bacteria, yeast, plants, antigen presenting cells (e.g., macrophage,

10 monocytes, dendritic cells, B-cells, T-cells, stem cells, and progenitor cells thereof), primary cells, immortalized cells, tumor-derived cells.

In a further aspect, the present invention includes compositions for generating an immunological response, where the composition typically comprises at least one of the expression cassettes of the present invention and may, for example, contain

15 combinations of expression cassettes (such as one or more expression cassettes carrying a Gag-polypeptide-encoding polynucleotide and one or more expression cassettes carrying an Env-polypeptide-encoding polynucleotide). Such compositions may further contain an adjuvant or adjuvants. The compositions may also contain one or more Gag-containing polypeptides and/or one or more Env-containing

20 polypeptides. The Gag-containing polypeptides and/or Env-containing polypeptides may correspond to the polypeptides encoded by the expression cassette(s) in the composition, or, the Gag-containing polypeptides and/or Env-containing polypeptides may be different from those encoded by the expression cassettes. An example of the polynucleotide in the expression cassette encoding the same polypeptide as is being

25 provided in the composition is as follows: the polynucleotide in the expression cassette encodes the Gag-polypeptide of Figure 1 (SEQ ID NO:3), and the polypeptide is the polypeptide encoded by the sequence shown in Figure 1 (SEQ ID NO:17). An example of the polynucleotide in the expression cassette encoding a different polypeptide as is being provided in the composition is as follows: an

30 expression cassette having a polynucleotide encoding a Gag-polymerase polypeptide, and the polypeptide provided in the composition may be a Gag and/or Gag-protease polypeptide. In compositions containing both expression cassettes (or

polynucleotides of the present invention) and polypeptides, the Env and Gag expression cassettes of the present invention can be mixed and/or matched with Env-containing and Gag-containing polypeptides described herein.

5 In another aspect the present invention includes methods of immunization of a subject. In the method any of the above described compositions are into the subject under conditions that are compatible with expression of the expression cassette in the subject. In one embodiment, the expression cassettes (or polynucleotides of the present invention) can be introduced using a gene delivery vector. The gene delivery vector can, for example, be a non-viral vector or a viral vector. Exemplary viral
10 vectors include, but are not limited to Sindbis-virus derived vectors, retroviral vectors, and lentiviral vectors. Compositions useful for generating an immunological response can also be delivered using a particulate carrier. Further, such compositions can be coated on, for example, gold or tungsten particles and the coated particles delivered to the subject using, for example, a gene gun. The compositions can also be formulated
15 as liposomes. In one embodiment of this method, the subject is a mammal and can, for example, be a human.

In a further aspect, the invention includes methods of generating an immune response in a subject, wherein the expression cassettes or polynucleotides of the present invention are expressed in a suitable cell to provide for the expression of the
20 Env- and/or Gag-containing polypeptides encoded by the polynucleotides of the present invention. The polypeptide(s) are then isolated (e.g., substantially purified) and administered to the subject in an amount sufficient to elicit an immune response.

The invention further includes methods of generating an immune response in a subject, where cells of a subject are transfected with any of the above-described
25 expression cassettes or polynucleotides of the present invention, under conditions that permit the expression of a selected polynucleotide and production of a polypeptide of interest (e.g., encoded by any expression cassette of the present invention). By this method an immunological response to the polypeptide is elicited in the subject. Transfection of the cells may be performed *ex vivo* and the transfected cells are
30 reintroduced into the subject. Alternately, or in addition, the cells may be transfected *in vivo* in the subject. The immune response may be humoral and/or cell-mediated (cellular). In a further embodiment, this method may also include administration of

an Env- and/or Gag-containing polypeptide before, concurrently with, and/or after introduction of the expression cassette into the subject.

Further embodiments of the present invention include purified polynucleotides. Exemplary polynucleotide sequences encoding *Gag*-containing polypeptides include, but are not limited to, the following polynucleotides:

5 nucleotides 844-903 of Figure 1 (SEQ ID NO:1) (a *Gag* major homology region); nucleotides 841-900 of Figure 2 (SEQ ID NO:2) (a *Gag* major homology region); the sequence presented as Figure 1 (SEQ ID NO:3); and the sequence presented as Figure 2 (SEQ ID NO:4). Exemplary polynucleotide sequences encoding *Env*-containing

10 polypeptides include, but are not limited to, the following polynucleotides: nucleotides 1213-1353 of Figure 3 (SEQ ID NO:5) (an *Env* common region); nucleotides 82-1512 of Figure 3 (SEQ ID NO:6) (a gp120 polypeptide); nucleotides 82-2025 of Figure 3 (SEQ ID NO:7) (a gp140 polypeptide); nucleotides 82-2547 of Figure 3 (SEQ ID NO:8) (a gp160 polypeptide); nucleotides 1-2547 of Figure 3 (SEQ

15 ID NO:9) (a gp160 polypeptide with signal sequence); nucleotides 1513-2547 of Figure 3 (SEQ ID NO:10) (a gp41 polypeptide); nucleotides 1210-1353 of Figure 4 (SEQ ID NO:11) (an *Env* common region); nucleotides 73-1509 of Figure 4 (SEQ ID NO:12) (a gp120 polypeptide); nucleotides 73-2022 of Figure 4 (SEQ ID NO:13) (a gp140 polypeptide); nucleotides 73-2565 of Figure 4 (SEQ ID NO:14) (a gp160

20 polypeptide); nucleotides 1-2565 of Figure 4 (SEQ ID NO:15) (a gp160 polypeptide with signal sequence); and nucleotides 1510-2565 of Figure 4 (SEQ ID NO:16) (a gp41 polypeptide). The polynucleotide sequence encoding the *Gag*-containing and *Env*-containing polypeptides of the present invention typically have at least about 85%, preferably about 90%, more preferably about 95%, and most preferably about

25 98% sequence identity to the sequences taught herein.

The polynucleotides of the present invention can be produced by recombinant techniques, synthetic techniques, or combinations thereof.

These and other embodiments of the present invention will readily occur to those of ordinary skill in the art in view of the disclosure herein.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 (SEQ ID NO:3) shows the nucleotide sequence of a polynucleotide encoding a synthetic Gag polypeptide. The nucleotide sequence shown was obtained by modifying type C strain AF110965 and include further modifications of INS.

5 Figure 2 (SEQ ID NO: 4) shows the nucleotide sequence of a polynucleotide encoding a synthetic Gag polypeptide. The nucleotide sequence shown was obtained by modifying type C strain AF110967 and include further modifications of INS.

10 Figure 3 (SEQ ID NO:9) shows the nucleotide sequence of a polynucleotide encoding a synthetic Env polypeptide. The nucleotide sequence depicts gp160 (including a signal peptide) and was obtained by modifying type C strain AF110968. The arrows indicate the positions of various regions of the polynucleotide, including the sequence encoding a signal peptide (nucleotides 1-81) (SEQ ID NO:18), a gp120 polypeptide (nucleotides 82-1512) (SEQ ID NO:6), a gp41 polypeptide (nucleotides 1513-2547) (SEQ ID NO:10), a gp140 polypeptide (nucleotides 82-2025) (SEQ ID
15 NO:7) and a gp160 polypeptide (nucleotides 82-2547) (SEQ ID NO:8). The codons encoding the signal peptide are modified (as described herein) from the native HIV-1 signal sequence.

20 Figure 4 (SEQ ID NO:15) shows the nucleotide sequence of a polynucleotide encoding a synthetic Env polypeptide. The nucleotide sequence depicts gp160 (including a signal peptide) and was obtained by modifying type C strain AF110975. The arrows indicate the positions of various regions of the polynucleotide, including the sequence encoding a signal peptide (nucleotides 1-72) (SEQ ID NO:19), a gp120 polypeptide (nucleotides 73-1509) (SEQ ID NO:12), a gp41 polypeptide (nucleotides 1510-2565) (SEQ ID NO:16), a gp140 polypeptide (nucleotides 73-2022) (SEQ ID
25 NO:13), and a gp160 polypeptide (nucleotides 73-2565) (SEQ ID NO:14). The codons encoding the signal peptide are modified (as described herein) from the native HIV-1 signal sequence.

30 Figure 5 shows the location of some remaining INS in synthetic Gag sequences derived from AF110965. The changes made to these sequences are boxed in the Figures. The top line depicts a codon optimized sequence of Gag polypeptides from the indicated strains (SEQ ID NO:20). The nucleotide(s) appearing below the

line in the boxed region(s) depicts changes made to remove further INS and correspond to the sequence depicted in Figure 1 (SEQ ID NO:3).

Figure 6 shows the location of some remaining INS in synthetic Gag sequences derived from AF110968. The changes made to these sequences are boxed in the Figures. The top line depicts a codon optimized sequence of Gag polypeptides from the indicated strains (SEQ ID NO:21). The nucleotide(s) appearing below the line in the boxed region(s) depicts changes made to remove further INS and correspond to the sequence depicted in Figure 2 (SEQ ID NO:4).

10 DETAILED DESCRIPTION OF THE INVENTION

The practice of the present invention will employ, unless otherwise indicated, conventional methods of chemistry, biochemistry, molecular biology, immunology and pharmacology, within the skill of the art. Such techniques are explained fully in the literature. See, e.g., *Remington's Pharmaceutical Sciences*, 18th Edition (Easton, Pennsylvania: Mack Publishing Company, 1990); *Methods In Enzymology* (S. Colowick and N. Kaplan, eds., Academic Press, Inc.); and *Handbook of Experimental Immunology*, Vols. I-IV (D.M. Weir and C.C. Blackwell, eds., 1986, Blackwell Scientific Publications); Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); *Short Protocols in Molecular Biology*, 4th ed. (Ausubel et al. eds., 1999, John Wiley & Sons); *Molecular Biology Techniques: An Intensive Laboratory Course*, (Ream et al., eds., 1998, Academic Press); *PCR (Introduction to Biotechniques Series)*, 2nd ed. (Newton & Graham eds., 1997, Springer Verlag).

As used in this specification and the appended claims, the singular forms "a," "an" and "the" include plural references unless the content clearly dictates otherwise. Thus, for example, reference to "an antigen" includes a mixture of two or more such agents.

1. DEFINITIONS

In describing the present invention, the following terms will be employed, and are intended to be defined as indicated below.

“Synthetic” sequences, as used herein, refers to Gag-encoding polynucleotides whose expression has been optimized as described herein, for example, by codon substitution and inactivation of inhibitory sequences. “Wild-type” or “native” sequences, as used herein, refers to polypeptide encoding sequences that are essentially as they are found in nature, e.g., Gag and/or Env encoding sequences as found in Type C isolates, e.g., AF110965, AF110967, AF110968 or AF110975.

As used herein, the term “virus-like particle” or “VLP” refers to a nonreplicating, viral shell, derived from any of several viruses discussed further below. VLPs are generally composed of one or more viral proteins, such as, but not limited to those proteins referred to as capsid, coat, shell, surface and/or envelope proteins, or particle-forming polypeptides derived from these proteins. VLPs can form spontaneously upon recombinant expression of the protein in an appropriate expression system. Methods for producing particular VLPs are known in the art and discussed more fully below. The presence of VLPs following recombinant expression of viral proteins can be detected using conventional techniques known in the art, such as by electron microscopy, X-ray crystallography, and the like. See, e.g., Baker et al., *Biophys. J.* (1991) 60:1445-1456; Hagensee et al., *J. Virol.* (1994) 68:4503-4505. For example, VLPs can be isolated by density gradient centrifugation and/or identified by characteristic density banding. Alternatively, cryoelectron microscopy can be performed on vitrified aqueous samples of the VLP preparation in question, and images recorded under appropriate exposure conditions.

By “particle-forming polypeptide” derived from a particular viral protein is meant a full-length or near full-length viral protein, as well as a fragment thereof, or a viral protein with internal deletions, which has the ability to form VLPs under conditions that favor VLP formation. Accordingly, the polypeptide may comprise the full-length sequence, fragments, truncated and partial sequences, as well as analogs and precursor forms of the reference molecule. The term therefore intends deletions, additions and substitutions to the sequence, so long as the polypeptide retains the ability to form a VLP. Thus, the term includes natural variations of the specified polypeptide since variations in coat proteins often occur between viral isolates. The term also includes deletions, additions and substitutions that do not naturally occur in the reference protein, so long as the protein retains the ability to form a VLP.

Preferred substitutions are those which are conservative in nature, i.e., those substitutions that take place within a family of amino acids that are related in their side chains. Specifically, amino acids are generally divided into four families: (1) acidic -- aspartate and glutamate; (2) basic -- lysine, arginine, histidine; (3) non-polar
5 -- alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar -- glycine, asparagine, glutamine, cystine, serine, threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified as aromatic amino acids.

An "antigen" refers to a molecule containing one or more epitopes (either
10 linear, conformational or both) that will stimulate a host's immune system to make a humoral and/or cellular antigen-specific response. The term is used interchangeably with the term "immunogen." Normally, a B-cell epitope will include at least about 5 amino acids but can be as small as 3-4 amino acids. A T-cell epitope, such as a CTL epitope, will include at least about 7-9 amino acids, and a helper T-cell epitope at least
15 about 12-20 amino acids. Normally, an epitope will include between about 7 and 15 amino acids, such as, 9, 10, 12 or 15 amino acids. The term "antigen" denotes both subunit antigens, (i.e., antigens which are separate and discrete from a whole organism with which the antigen is associated in nature), as well as, killed, attenuated or inactivated bacteria, viruses, fungi, parasites or other microbes. Antibodies such as
20 anti-idiotypic antibodies, or fragments thereof, and synthetic peptide mimotopes, which can mimic an antigen or antigenic determinant, are also captured under the definition of antigen as used herein. Similarly, an oligonucleotide or polynucleotide which expresses an antigen or antigenic determinant *in vivo*, such as in gene therapy and DNA immunization applications, is also included in the definition of antigen
25 herein.

For purposes of the present invention, antigens can be derived from any of several known viruses, bacteria, parasites and fungi, as described more fully below. The term also intends any of the various tumor antigens. Furthermore, for purposes of the present invention, an "antigen" refers to a protein which includes modifications,
30 such as deletions, additions and substitutions (generally conservative in nature), to the native sequence, so long as the protein maintains the ability to elicit an immunological response, as defined herein. These modifications may be deliberate,

as through site-directed mutagenesis, or may be accidental, such as through mutations of hosts which produce the antigens.

An "immunological response" to an antigen or composition is the development in a subject of a humoral and/or a cellular immune response to an antigen present in the composition of interest. For purposes of the present invention, a "humoral immune response" refers to an immune response mediated by antibody molecules, while a "cellular immune response" is one mediated by T-lymphocytes and/or other white blood cells. One important aspect of cellular immunity involves an antigen-specific response by cytolytic T-cells ("CTL"s). CTLs have specificity for peptide antigens that are presented in association with proteins encoded by the major histocompatibility complex (MHC) and expressed on the surfaces of cells. CTLs help induce and promote the destruction of intracellular microbes, or the lysis of cells infected with such microbes. Another aspect of cellular immunity involves an antigen-specific response by helper T-cells. Helper T-cells act to help stimulate the function, and focus the activity of, nonspecific effector cells against cells displaying peptide antigens in association with MHC molecules on their surface. A "cellular immune response" also refers to the production of cytokines, chemokines and other such molecules produced by activated T-cells and/or other white blood cells, including those derived from CD4+ and CD8+ T-cells.

A composition or vaccine that elicits a cellular immune response may serve to sensitize a vertebrate subject by the presentation of antigen in association with MHC molecules at the cell surface. The cell-mediated immune response is directed at, or near, cells presenting antigen at their surface. In addition, antigen-specific T-lymphocytes can be generated to allow for the future protection of an immunized host.

The ability of a particular antigen to stimulate a cell-mediated immunological response may be determined by a number of assays, such as by lymphoproliferation (lymphocyte activation) assays, CTL cytotoxic cell assays, or by assaying for T-lymphocytes specific for the antigen in a sensitized subject. Such assays are well known in the art. See, e.g., Erickson et al., *J. Immunol.* (1993) 151:4189-4199; Doe et al., *Eur. J. Immunol.* (1994) 24:2369-2376. Recent methods of measuring cell-mediated immune response include measurement of intracellular cytokines or cytokine secretion by T-cell populations, or by measurement of epitope specific T-

cells (e.g., by the tetramer technique)(reviewed by McMichael, A.J., and O'Callaghan, C.A., *J. Exp. Med.* **187**(9)1367-1371, 1998; McHeyzer-Williams, M.G., et al, *Immunol. Rev.* **150**:5-21, 1996; Lalvani, A., et al, *J. Exp. Med.* **186**:859-865, 1997).

Thus, an immunological response as used herein may be one which stimulates
5 the production of CTLs, and/or the production or activation of helper T- cells. The antigen of interest may also elicit an antibody-mediated immune response. Hence, an immunological response may include one or more of the following effects: the production of antibodies by B-cells; and/or the activation of suppressor T-cells and/or $\gamma\delta$ T-cells directed specifically to an antigen or antigens present in the composition or
10 vaccine of interest. These responses may serve to neutralize infectivity, and/or mediate antibody-complement, or antibody dependent cell cytotoxicity (ADCC) to provide protection to an immunized host. Such responses can be determined using standard immunoassays and neutralization assays, well known in the art.

An "immunogenic composition" is a composition that comprises an antigenic
15 molecule where administration of the composition to a subject results in the development in the subject of a humoral and/or a cellular immune response to the antigenic molecule of interest. The immunogenic composition can be introduced directly into a recipient subject, such as by injection, inhalation, oral, intranasal and mucosal (e.g., intra-rectally or intra-vaginally) administration.

By "subunit vaccine" is meant a vaccine composition which includes one or
20 more selected antigens but not all antigens, derived from or homologous to, an antigen from a pathogen of interest such as from a virus, bacterium, parasite or fungus. Such a composition is substantially free of intact pathogen cells or pathogenic particles, or the lysate of such cells or particles. Thus, a "subunit vaccine"
25 can be prepared from at least partially purified (preferably substantially purified) immunogenic polypeptides from the pathogen, or analogs thereof. The method of obtaining an antigen included in the subunit vaccine can thus include standard purification techniques, recombinant production, or synthetic production.

"Substantially purified" general refers to isolation of a substance (compound,
30 polynucleotide, protein, polypeptide, polypeptide composition) such that the substance comprises the majority percent of the sample in which it resides. Typically in a sample a substantially purified component comprises 50%, preferably 80%-85%,

more preferably 90-95% of the sample. Techniques for purifying polynucleotides and polypeptides of interest are well-known in the art and include, for example, ion-exchange chromatography, affinity chromatography and sedimentation according to density.

5 A "coding sequence" or a sequence which "encodes" a selected polypeptide, is a nucleic acid molecule which is transcribed (in the case of DNA) and translated (in the case of mRNA) into a polypeptide *in vivo* when placed under the control of appropriate regulatory sequences (or "control elements"). The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a
10 translation stop codon at the 3' (carboxy) terminus. A coding sequence can include, but is not limited to, cDNA from viral, procaryotic or eucaryotic mRNA, genomic DNA sequences from viral or procaryotic DNA, and even synthetic DNA sequences. A transcription termination sequence may be located 3' to the coding sequence.

 Typical "control elements", include, but are not limited to, transcription
15 promoters, transcription enhancer elements, transcription termination signals, polyadenylation sequences (located 3' to the translation stop codon), sequences for optimization of initiation of translation (located 5' to the coding sequence), and translation termination sequences.

 A "nucleic acid" molecule can include, but is not limited to, procaryotic
20 sequences, eucaryotic mRNA, cDNA from eucaryotic mRNA, genomic DNA sequences from eucaryotic (e.g., mammalian) DNA, and even synthetic DNA sequences. The term also captures sequences that include any of the known base analogs of DNA and RNA.

 "Operably linked" refers to an arrangement of elements wherein the
25 components so described are configured so as to perform their usual function. Thus, a given promoter operably linked to a coding sequence is capable of effecting the expression of the coding sequence when the proper enzymes are present. The promoter need not be contiguous with the coding sequence, so long as it functions to direct the expression thereof. Thus, for example, intervening untranslated yet
30 transcribed sequences can be present between the promoter sequence and the coding sequence and the promoter sequence can still be considered "operably linked" to the coding sequence.

"Recombinant" as used herein to describe a nucleic acid molecule means a polynucleotide of genomic, cDNA, semisynthetic, or synthetic origin which, by virtue of its origin or manipulation: (1) is not associated with all or a portion of the polynucleotide with which it is associated in nature; and/or (2) is linked to a polynucleotide other than that to which it is linked in nature. The term "recombinant" as used with respect to a protein or polypeptide means a polypeptide produced by expression of a recombinant polynucleotide. "Recombinant host cells," "host cells," "cells," "cell lines," "cell cultures," and other such terms denoting procaryotic microorganisms or eucaryotic cell lines cultured as unicellular entities, are used interchangeably, and refer to cells which can be, or have been, used as recipients for recombinant vectors or other transfer DNA, and include the progeny of the original cell which has been transfected. It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement to the original parent, due to accidental or deliberate mutation. Progeny of the parental cell which are sufficiently similar to the parent to be characterized by the relevant property, such as the presence of a nucleotide sequence encoding a desired peptide, are included in the progeny intended by this definition, and are covered by the above terms.

Techniques for determining amino acid sequence "similarity" are well known in the art. In general, "similarity" means the exact amino acid to amino acid comparison of two or more polypeptides at the appropriate place, where amino acids are identical or possess similar chemical and/or physical properties such as charge or hydrophobicity. A so-termed "percent similarity" then can be determined between the compared polypeptide sequences. Techniques for determining nucleic acid and amino acid sequence identity also are well known in the art and include determining the nucleotide sequence of the mRNA for that gene (usually via a cDNA intermediate) and determining the amino acid sequence encoded thereby, and comparing this to a second amino acid sequence. In general, "identity" refers to an exact nucleotide to nucleotide or amino acid to amino acid correspondence of two polynucleotides or polypeptide sequences, respectively.

Two or more polynucleotide sequences can be compared by determining their "percent identity." Two or more amino acid sequences likewise can be compared by

determining their "percent identity." The percent identity of two sequences, whether nucleic acid or peptide sequences, is generally described as the number of exact matches between two aligned sequences divided by the length of the shorter sequence and multiplied by 100. An approximate alignment for nucleic acid sequences is provided by the local homology algorithm of Smith and Waterman, *Advances in Applied Mathematics* 2:482-489 (1981). This algorithm can be extended to use with peptide sequences using the scoring matrix developed by Dayhoff, *Atlas of Protein Sequences and Structure*, M.O. Dayhoff ed., 5 suppl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., USA, and normalized by Gribskov, *Nucl. Acids Res.* 14(6):6745-6763 (1986). An implementation of this algorithm for nucleic acid and peptide sequences is provided by the Genetics Computer Group (Madison, WI) in their BestFit utility application. The default parameters for this method are described in the Wisconsin Sequence Analysis Package Program Manual, Version 8 (1995) (available from Genetics Computer Group, Madison, WI). Other equally suitable programs for calculating the percent identity or similarity between sequences are generally known in the art.

For example, percent identity of a particular nucleotide sequence to a reference sequence can be determined using the homology algorithm of Smith and Waterman with a default scoring table and a gap penalty of six nucleotide positions. Another method of establishing percent identity in the context of the present invention is to use the MPSRCH package of programs copyrighted by the University of Edinburgh, developed by John F. Collins and Shane S. Sturrok, and distributed by IntelliGenetics, Inc. (Mountain View, CA). From this suite of packages, the Smith-Waterman algorithm can be employed where default parameters are used for the scoring table (for example, gap open penalty of 12, gap extension penalty of one, and a gap of six). From the data generated, the "Match" value reflects "sequence identity." Other suitable programs for calculating the percent identity or similarity between sequences are generally known in the art, such as the alignment program BLAST, which can also be used with default parameters. For example, BLASTN and BLASTP can be used with the following default parameters: genetic code = standard; filter = none; strand = both; cutoff = 60; expect = 10; Matrix = BLOSUM62; Descriptions = 50 sequences; sort by = HIGH SCORE; Databases = non-redundant, GenBank + EMBL + DDBJ +

PDB + GenBank CDS translations + Swiss protein + Spupdate + PIR. Details of these programs can be found at the following internet address:

<http://www.ncbi.nlm.gov/cgi-bin/BLAST>.

One of skill in the art can readily determine the proper search parameters to use for a given sequence in the above programs. For example, the search parameters may vary based on the size of the sequence in question. Thus, for example, a representative embodiment of the present invention would include an isolated polynucleotide having X contiguous nucleotides, wherein (i) the X contiguous nucleotides have at least about 50% identity to Y contiguous nucleotides derived from any of the sequences described herein, (ii) X equals Y, and (iii) X is greater than or equal to 6 nucleotides and up to 5000 nucleotides, preferably greater than or equal to 8 nucleotides and up to 5000 nucleotides, more preferably 10-12 nucleotides and up to 5000 nucleotides, and even more preferably 15-20 nucleotides, up to the number of nucleotides present in the full-length sequences described herein (e.g., see the Sequence Listing and claims), including all integer values falling within the above-described ranges.

The synthetic expression cassettes (and purified polynucleotides) of the present invention include related polynucleotide sequences having about 80% to 100%, greater than 80-85%, preferably greater than 90-92%, more preferably greater than 95%, and most preferably greater than 98% sequence (including all integer values falling within these described ranges) identity to the synthetic expression cassette sequences disclosed herein (for example, to the claimed sequences or other sequences of the present invention) when the sequences of the present invention are used as the query sequence.

Two nucleic acid fragments are considered to "selectively hybridize" as described herein. The degree of sequence identity between two nucleic acid molecules affects the efficiency and strength of hybridization events between such molecules. A partially identical nucleic acid sequence will at least partially inhibit a completely identical sequence from hybridizing to a target molecule. Inhibition of hybridization of the completely identical sequence can be assessed using hybridization assays that are well known in the art (e.g., Southern blot, Northern blot, solution hybridization, or the like, see Sambrook, et al., *supra* or Ausubel et al.,

supra). Such assays can be conducted using varying degrees of selectivity, for example, using conditions varying from low to high stringency. If conditions of low stringency are employed, the absence of non-specific binding can be assessed using a secondary probe that lacks even a partial degree of sequence identity (for example, a probe having less than about 30% sequence identity with the target molecule), such that, in the absence of non-specific binding events, the secondary probe will not hybridize to the target.

When utilizing a hybridization-based detection system, a nucleic acid probe is chosen that is complementary to a target nucleic acid sequence, and then by selection of appropriate conditions the probe and the target sequence "selectively hybridize," or bind, to each other to form a hybrid molecule. A nucleic acid molecule that is capable of hybridizing selectively to a target sequence under "moderately stringent" typically hybridizes under conditions that allow detection of a target nucleic acid sequence of at least about 10-14 nucleotides in length having at least approximately 70% sequence identity with the sequence of the selected nucleic acid probe. Stringent hybridization conditions typically allow detection of target nucleic acid sequences of at least about 10-14 nucleotides in length having a sequence identity of greater than about 90-95% with the sequence of the selected nucleic acid probe. Hybridization conditions useful for probe/target hybridization where the probe and target have a specific degree of sequence identity, can be determined as is known in the art (see, for example, Nucleic Acid Hybridization: A Practical Approach, editors B.D. Hames and S.J. Higgins, (1985) Oxford; Washington, DC; IRL Press).

With respect to stringency conditions for hybridization, it is well known in the art that numerous equivalent conditions can be employed to establish a particular stringency by varying, for example, the following factors: the length and nature of probe and target sequences, base composition of the various sequences, concentrations of salts and other hybridization solution components, the presence or absence of blocking agents in the hybridization solutions (e.g., formamide, dextran sulfate, and polyethylene glycol), hybridization reaction temperature and time parameters, as well as, varying wash conditions. The selection of a particular set of hybridization conditions is selected following standard methods in the art (see, for example, Sambrook, et al., *supra* or Ausubel et al., *supra*).

A first polynucleotide is "derived from" second polynucleotide if it has the same or substantially the same basepair sequence as a region of the second polynucleotide, its cDNA, complements thereof, or if it displays sequence identity as described above.

5 A first polypeptide is "derived from" a second polypeptide if it is (i) encoded by a first polynucleotide derived from a second polynucleotide, or (ii) displays sequence identity to the second polypeptides as described above.

10 Generally, a viral polypeptide is "derived from" a particular polypeptide of a virus (viral polypeptide) if it is (i) encoded by an open reading frame of a polynucleotide of that virus (viral polynucleotide), or (ii) displays sequence identity to polypeptides of that virus as described above.

15 "Encoded by" refers to a nucleic acid sequence which codes for a polypeptide sequence, wherein the polypeptide sequence or a portion thereof contains an amino acid sequence of at least 3 to 5 amino acids, more preferably at least 8 to 10 amino acids, and even more preferably at least 15 to 20 amino acids from a polypeptide encoded by the nucleic acid sequence. Also encompassed are polypeptide sequences which are immunologically identifiable with a polypeptide encoded by the sequence.

20 "Purified polynucleotide" refers to a polynucleotide of interest or fragment thereof which is essentially free, e.g., contains less than about 50%, preferably less than about 70%, and more preferably less than about 90%, of the protein with which the polynucleotide is naturally associated. Techniques for purifying polynucleotides of interest are well-known in the art and include, for example, disruption of the cell containing the polynucleotide with a chaotropic agent and separation of the polynucleotide(s) and proteins by ion-exchange chromatography, affinity
25 chromatography and sedimentation according to density.

30 By "nucleic acid immunization" is meant the introduction of a nucleic acid molecule encoding one or more selected antigens into a host cell, for the *in vivo* expression of an antigen, antigens, an epitope, or epitopes. The nucleic acid molecule can be introduced directly into a recipient subject, such as by injection, inhalation, oral, intranasal and mucosal administration, or the like, or can be introduced *ex vivo*, into cells which have been removed from the host. In the latter case, the transformed

cells are reintroduced into the subject where an immune response can be mounted against the antigen encoded by the nucleic acid molecule.

"Gene transfer" or "gene delivery" refers to methods or systems for reliably inserting DNA of interest into a host cell. Such methods can result in transient
5 expression of non-integrated transferred DNA, extrachromosomal replication and expression of transferred replicons (e.g., episomes), or integration of transferred genetic material into the genomic DNA of host cells. Gene delivery expression vectors include, but are not limited to, vectors derived from alphaviruses, pox viruses and vaccinia viruses. When used for immunization, such gene delivery expression
10 vectors may be referred to as vaccines or vaccine vectors.

"T lymphocytes" or "T cells" are non-antibody producing lymphocytes that constitute a part of the cell-mediated arm of the immune system. T cells arise from immature lymphocytes that migrate from the bone marrow to the thymus, where they undergo a maturation process under the direction of thymic hormones. Here, the
15 mature lymphocytes rapidly divide increasing to very large numbers. The maturing T cells become immunocompetent based on their ability to recognize and bind a specific antigen. Activation of immunocompetent T cells is triggered when an antigen binds to the lymphocyte's surface receptors.

The term "transfection" is used to refer to the uptake of foreign DNA by a cell.
20 A cell has been "transfected" when exogenous DNA has been introduced inside the cell membrane. A number of transfection techniques are generally known in the art. See, e.g., Graham et al. (1973) *Virology*, 52:456, Sambrook et al. (1989) *Molecular Cloning, a laboratory manual*, Cold Spring Harbor Laboratories, New York, Davis et al. (1986) *Basic Methods in Molecular Biology*, Elsevier, and Chu et al. (1981) *Gene*
25 13:197. Such techniques can be used to introduce one or more exogenous DNA moieties into suitable host cells. The term refers to both stable and transient uptake of the genetic material, and includes uptake of peptide- or antibody-linked DNAs.

A "vector" is capable of transferring gene sequences to target cells (e.g., viral vectors, non-viral vectors, particulate carriers, and liposomes). Typically, "vector
30 construct," "expression vector," and "gene transfer vector," mean any nucleic acid construct capable of directing the expression of a gene of interest and which can

transfer gene sequences to target cells. Thus, the term includes cloning and expression vehicles, as well as viral vectors.

Transfer of a "suicide gene" (e.g., a drug-susceptibility gene) to a target cell renders the cell sensitive to compounds or compositions that are relatively nontoxic to normal cells. Moolten, F.L. (1994) *Cancer Gene Ther.* 1:279-287. Examples of suicide genes are thymidine kinase of herpes simplex virus (HSV-tk), cytochrome P450 (Manome et al. (1996) *Gene Therapy* 3:513-520), human deoxycytidine kinase (Manome et al. (1996) *Nature Medicine* 2(5):567-573) and the bacterial enzyme cytosine deaminase (Dong et al. (1996) *Human Gene Therapy* 7:713-720). Cells which express these genes are rendered sensitive to the effects of the relatively nontoxic prodrugs ganciclovir (HSV-tk), cyclophosphamide (cytochrome P450 2B1), cytosine arabinoside (human deoxycytidine kinase) or 5-fluorocytosine (bacterial cytosine deaminase). Culver et al. (1992) *Science* 256:1550-1552, Huber et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:8302-8306.

A "selectable marker" or "reporter marker" refers to a nucleotide sequence included in a gene transfer vector that has no therapeutic activity, but rather is included to allow for simpler preparation, manufacturing, characterization or testing of the gene transfer vector.

A "specific binding agent" refers to a member of a specific binding pair of molecules wherein one of the molecules specifically binds to the second molecule through chemical and/or physical means. One example of a specific binding agent is an antibody directed against a selected antigen.

By "subject" is meant any member of the subphylum chordata, including, without limitation, humans and other primates, including non-human primates such as chimpanzees and other apes and monkey species; farm animals such as cattle, sheep, pigs, goats and horses; domestic mammals such as dogs and cats; laboratory animals including rodents such as mice, rats and guinea pigs; birds, including domestic, wild and game birds such as chickens, turkeys and other gallinaceous birds, ducks, geese, and the like. The term does not denote a particular age. Thus, both adult and newborn individuals are intended to be covered. The system described above is intended for use in any of the above vertebrate species, since the immune systems of all of these vertebrates operate similarly.

By "pharmaceutically acceptable" or "pharmacologically acceptable" is meant a material which is not biologically or otherwise undesirable, i.e., the material may be administered to an individual in a formulation or composition without causing any undesirable biological effects or interacting in a deleterious manner with any of the components of the composition in which it is contained.

By "physiological pH" or a "pH in the physiological range" is meant a pH in the range of approximately 7.2 to 8.0 inclusive, more typically in the range of approximately 7.2 to 7.6 inclusive.

As used herein, "treatment" refers to any of (i) the prevention of infection or reinfection, as in a traditional vaccine, (ii) the reduction or elimination of symptoms, and (iii) the substantial or complete elimination of the pathogen in question. Treatment may be effected prophylactically (prior to infection) or therapeutically (following infection).

"Lentiviral vector", and "recombinant lentiviral vector" refer to a nucleic acid construct which carries, and within certain embodiments, is capable of directing the expression of a nucleic acid molecule of interest. The lentiviral vector include at least one transcriptional promoter/enhancer or locus defining element(s), or other elements which control gene expression by other means such as alternate splicing, nuclear RNA export, post-translational modification of messenger, or post-transcriptional modification of protein. Such vector constructs must also include a packaging signal, long terminal repeats (LTRS) or portion thereof, and positive and negative strand primer binding sites appropriate to the retrovirus used (if these are not already present in the retroviral vector). Optionally, the recombinant lentiviral vector may also include a signal which directs polyadenylation, selectable markers such as Neo, TK, hygromycin, phleomycin, histidinol, or DHFR, as well as one or more restriction sites and a translation termination sequence. By way of example, such vectors typically include a 5' LTR, a tRNA binding site, a packaging signal, an origin of second strand DNA synthesis, and a 3'LTR or a portion thereof

"Lentiviral vector particle" as utilized within the present invention refers to a lentivirus which carries at least one gene of interest. The retrovirus may also contain a selectable marker. The recombinant lentivirus is capable of reverse transcribing its genetic material (RNA) into DNA and incorporating this genetic material into a host

cell's DNA upon infection. Lentiviral vector particles may have a lentiviral envelope, a non-lentiviral envelope (e.g., an amphi or VSV-G envelope), or a chimeric envelope.

"Nucleic acid expression vector" or "Expression cassette" refers to an assembly which is capable of directing the expression of a sequence or gene of interest. The nucleic acid expression vector includes a promoter which is operably linked to the sequences or gene(s) of interest. Other control elements may be present as well. Expression cassettes described herein may be contained within a plasmid construct. In addition to the components of the expression cassette, the plasmid construct may also include a bacterial origin of replication, one or more selectable markers, a signal which allows the plasmid construct to exist as single-stranded DNA (e.g., a M13 origin of replication), a multiple cloning site, and a "mammalian" origin of replication (e.g., a SV40 or adenovirus origin of replication).

"Packaging cell" refers to a cell which contains those elements necessary for production of infectious recombinant retrovirus which are lacking in a recombinant retroviral vector. Typically, such packaging cells contain one or more expression cassettes which are capable of expressing proteins which encode *Gag*, *pol* and *env* proteins.

"Producer cell" or "vector producing cell" refers to a cell which contains all elements necessary for production of recombinant retroviral vector particles.

2. MODES OF CARRYING OUT THE INVENTION

Before describing the present invention in detail, it is to be understood that this invention is not limited to particular formulations or process parameters as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments of the invention only, and is not intended to be limiting.

Although a number of methods and materials similar or equivalent to those described herein can be used in the practice of the present invention, the preferred materials and methods are described herein.

2.1 SYNTHETIC EXPRESSION CASSETTES

2.1.1 MODIFICATION OF HIV-1-TYPE C GAG AND ENV NUCLEIC ACID CODING SEQUENCES

One aspect of the present invention is the generation of HIV-1 type C Gag and
5 Env protein coding sequences, and related sequences, having improved expression relative to the corresponding wild-type sequences.

2.1.1.1. MODIFICATION OF GAG NUCLEIC ACID CODING SEQUENCES

An exemplary embodiment of the present invention is illustrated herein by
10 modifying the Gag protein wild-type sequences obtained from the AF110965 and AF110967 strains of HIV-1, subtype C. (see, for example, Korber et al. (1998)*Human Retroviruses and Aids*, Los Alamos, New Mexico: Los Alamos National Laboratory; Novitsky et al. (1999) *J. Virol.* 73(5):4427-4432, for molecular cloning of various subtype C clones from Botswana). Gag sequence obtained from
15 other Type C HIV-1 variants may be manipulated in similar fashion following the teachings of the present specification. Such other variants include, but are not limited to, Gag protein encoding sequences obtained from the isolates of HIV-1 Type C, for example as described in Novitsky et al., (1999), *supra*; Myers et al., *infra*; Virology, 3rd Edition (W.K. Joklik ed. 1988); *Fundamental Virology*, 2nd Edition (B.N. Fields and D.M. Knipe, eds. 1991); *Virology*, 3rd Edition (Fields, BN, DM Knipe, PM
20 Howley, Editors, 1996, Lippincott-Raven, Philadelphia, PA and on the World Wide Web (Internet), for example at <http://hiv-web.lanl.gov/cgi-bin/hivDB3/public/wdb/ssampublic> and <http://hiv-web.lanl.gov>.

First, the HIV-1 codon usage pattern was modified so that the resulting nucleic
25 acid coding sequence was comparable to codon usage found in highly expressed human genes (Example 1). The HIV codon usage reflects a high content of the nucleotides A or T of the codon-triplet. The effect of the HIV-1 codon usage is a high AT content in the DNA sequence that results in a decreased translation ability and instability of the mRNA. In comparison, highly expressed human codons prefer the
30 nucleotides G or C. The Gag coding sequences were modified to be comparable to codon usage found in highly expressed human genes.

Second, there are inhibitory (or instability) elements (INS) located within the coding sequences of the Gag coding sequences. The RRE is a secondary RNA structure that interacts with the HIV encoded Rev-protein to overcome the expression down-regulating effects of the INS. To overcome the post-transcriptional activating mechanisms of RRE and Rev, the instability elements can be inactivated by introducing multiple point mutations that do not alter the reading frame of the encoded proteins. Subtype C Gag-encoding sequences having inactivated RRE sites are shown in Figures 1 (SEQ ID NO:3), 2 (SEQ ID NO:4), 5 (SEQ ID NO:20) and 6 (SEQ ID NO:26).

Modification of the Gag polypeptide coding sequences results in improved expression relative to the wild-type coding sequences in a number of mammalian cell lines (as well as other types of cell lines, including, but not limited to, insect cells). Further, expression of the sequences results in production of virus-like particles (VLPs) by these cell lines (see below).

2.1.1.2 MODIFICATION OF *ENV* NUCLEIC ACID CODING SEQUENCES

Similarly, the present invention also includes modified Env proteins. Wild-type Env sequences are obtained from the AF110968 and AF110975 strains of HIV-1, type C. (see, for example, Novitsky et al. (1999) *J. Virol.* 73(5):4427-4432, for molecular cloning of various subtype C clones from Botswana). Env sequence obtained from other Type C HIV-1 variants may be manipulated in similar fashion following the teachings of the present specification. Such other variants include, but are not limited to, Env protein encoding sequences obtained from the isolates of HIV-1 Type C, described above.

The codon usage pattern for Env was modified as described above for Gag so that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human genes. Experiments can be performed in support of the present invention to show that the synthetic Env sequences were capable of higher level of protein production relative to the native Env sequences.

Modification of the Env polypeptide coding sequences results in improved expression relative to the wild-type coding sequences in a number of mammalian cell lines (as well as other types of cell lines, including, but not limited to, insect cells).

Similar Env polypeptide coding sequences can be obtained, optimized and tested for improved expression from a variety of isolates, including those described above for Gag.

5 **2.1.2 FURTHER MODIFICATION OF SEQUENCES INCLUDING HIV-1 GAG
 NUCLEIC ACID CODING SEQUENCES**

Experiments can be performed to show that similar modifications of HIV-1 Gag-protease and Gag-polymerase sequences also result in improved expression of the polyproteins, as well as, the production of VLPs formed by polypeptides produced
10 from such modified coding sequences.

For the Gag-protease sequence, the changes in codon usage are typically restricted to the regions up to the -1 frameshift and starting again at the end of the Gag reading frame; however, regions within the frameshift translation region can be modified as well. Further, inhibitory (or instability) elements (INS) located within the
15 coding sequences of the Gag-protease polypeptide coding sequence can be altered as well.

For the Gag-polymerase sequence, the changes in codon usage can be similar to those for the Gag-protease sequence.

In addition to polyproteins containing HIV-related sequences, the Gag
20 encoding sequences of the present invention can be fused to other polypeptides (creating chimeric polypeptides) for which an immunogenic response is desired.

Further sequences useful in the practice of the present invention include, but are not limited to, sequences encoding further viral epitopes/antigens {including but not limited to, HCV antigens (e.g., E1, E2; Houghton, M., et al., U.S. Patent No. 5,714,596, issued February 3, 1998; Houghton, M., et al., U.S. Patent No. 5,712,088,
25 issued January 27, 1998; Houghton, M., et al., U.S. Patent No. 5,683,864, issued November 4, 1997; Weiner, A.J., et al., U.S. Patent No. 5,728,520, issued March 17, 1998; Weiner, A.J., et al., U.S. Patent No. 5,766,845, issued June 16, 1998; Weiner, A.J., et al., U.S. Patent No. 5,670,152, issued September 23, 1997), HIV antigens
30 (e.g., derived from *tat*, *rev*, *nef* and/or *env*); and sequences encoding tumor antigens/epitopes. Additional sequences are described below. Also, variations on the

orientation of the Gag and other coding sequences, relative to each other, are described below.

Gag, Gag-protease, and/or Gag-polymerase polypeptide coding sequences can be obtained from Type C HIV isolates, see, e.g., Myers et al. Los Alamos Database,
5 Los Alamos National Laboratory, Los Alamos, New Mexico (1992); Myers et al.,
Human Retroviruses and Aids, 1997, Los Alamos, New Mexico: Los Alamos
National Laboratory. Synthetic expression cassettes can be generated using such
coding sequences as starting material by following the teachings of the present
specification (e.g., see Example 1).

10 Further, the synthetic expression cassettes of the present invention include
related Gag polypeptide coding sequences having greater than 85%, preferably greater
than 90%, more preferably greater than 95%, and most preferably greater than 98%
sequence identity to the synthetic expression cassette sequences disclosed herein (for
example, Figures 1, 2, 5 and 6 (SEQ ID NOs:3, 4, 20 and 21).

15 The present invention also includes related Env polypeptide coding sequences
having greater than 85%, preferably greater than 90%, more preferably greater than
95%, and most preferably greater than 98% sequence identity to the sequences
disclosed herein (for example, Figures 3 and 4, SEQ ID NOs:5-17). Various coding
regions are indicated in Figures 3 and 4, for example in Figure 3 (AF110968),
20 nucleotides 1-81 (SEQ ID NO:18) encode a signal peptide, nucleotides 82-1512 (SEQ
ID NO:6) encode a gp120 polypeptide, nucleotides 1513 to 2547 (SEQ ID NO:10)
encode a gp41 polypeptide, nucleotides 82-2025 (SEQ ID NO:7) encode a gp140
polypeptide and nucleotides 82-2547 (SEQ ID NO:8) encode a gp160 polypeptide.

25 2.1.3 EXPRESSION OF SYNTHETIC SEQUENCES ENCODING HIV-1 GAG OR ENV AND RELATED POLYPEPTIDES

Synthetic Gag- and Env-encoding sequences (expression cassettes) of the
present invention can be cloned into a number of different expression vectors to
evaluate levels of expression and, in the case of Gag, production of VLPs. The
30 synthetic DNA fragments for Env and Gag can be cloned into eucaryotic expression
vectors, including, a transient expression vector, CMV-promoter-based mammalian

vectors, and a shuttle vector for use in baculovirus expression systems.

Corresponding wild-type sequences can also be cloned into the same vectors.

5 These vectors can then be transfected into a several different cell types, including a variety of mammalian cell lines, (293, RD, COS-7, and CHO, cell lines available, for example, from the A.T.C.C.). The cell lines are then cultured under appropriate conditions and the levels of p24 (Gag) or, gp160 or gp120 (Env) expression in supernatants can be evaluated (Example 2). Env polypeptides include, but are not limited to, for example, native gp160, oligomeric gp140, monomeric gp120 as well as modified sequences of these polypeptides. The results of these
10 assays demonstrate that expression of synthetic Env, Gag and Gag-protease encoding sequences are significantly higher than corresponding wild-type sequences.

 Further, Western Blot analysis can be used to show that cells containing the synthetic Gag or Env expression cassette produce the expected protein at higher per-cell concentrations than cells containing the native expression cassette. The Gag and
15 Env proteins can be seen in both cell lysates and supernatants. The levels of production are significantly higher in cell supernatants for cells transfected with the synthetic expression cassettes of the present invention.

 Fractionation of the supernatants from mammalian cells transfected with the synthetic Gag or Env expression cassette can be used to show that the cassettes
20 provide superior production of both Gag and Env proteins and, in the case of Gag, VLPs, relative to the wild-type sequences.

 Efficient expression of these Gag- and/or Env-containing polypeptides in mammalian cell lines provides the following benefits: the polypeptides are free of baculovirus contaminants; production by established methods approved by the FDA;
25 increased purity; greater yields (relative to native coding sequences); and a novel method of producing the Gag- and/or Env-containing polypeptides in CHO cells which is not feasible in the absence of the increased expression obtained using the constructs of the present invention. Exemplary Mammalian cell lines include, but are not limited to, BHK, VERO, HT1080, 293, 293T, RD, COS-7, CHO, Jurkat, HUT,
30 SUPT, C8166, MOLT4/clone8, MT-2, MT-4, H9, PM1, CEM, and CEMX174, such cell lines are available, for example, from the A.T.C.C.).

A synthetic Gag expression cassette of the present invention will also exhibit high levels of expression and VLP production when transfected into insect cells. Synthetic Env expression cassettes also demonstrate high levels of expression in insect cells. Further, in addition to a higher total protein yield, the final product from the synthetic polypeptides consistently contains lower amounts of contaminating baculovirus proteins than the final product from the native Gag or Env.

Further, synthetic Gag and Env expression cassettes of the present invention can also be introduced into yeast vectors which, in turn, can be transformed into and efficiently expressed by yeast cells (*Saccharomyces cerevisiae*; using vectors as described in Rosenberg, S. and Tekamp-Olson, P., U.S. Patent No. RE35,749, issued, March 17, 1998).

In addition to the mammalian and insect vectors, the synthetic expression cassettes of the present invention can be incorporated into a variety of expression vectors using selected expression control elements. Appropriate vectors and control elements for any given cell type can be selected by one having ordinary skill in the art in view of the teachings of the present specification and information known in the art about expression vectors.

For example, a synthetic Gag or Env expression cassette can be inserted into a vector which includes control elements operably linked to the desired coding sequence, which allow for the expression of the gene in a selected cell-type. For example, typical promoters for mammalian cell expression include the SV40 early promoter, a CMV promoter such as the CMV immediate early promoter (a CMV promoter can include intron A), RSV, HIV-Ltr, the mouse mammary tumor virus LTR promoter (MMLV-ltr), the adenovirus major late promoter (Ad MLP), and the herpes simplex virus promoter, among others. Other nonviral promoters, such as a promoter derived from the murine metallothionein gene, will also find use for mammalian expression. Typically, transcription termination and polyadenylation sequences will also be present, located 3' to the translation stop codon. Preferably, a sequence for optimization of initiation of translation, located 5' to the coding sequence, is also present. Examples of transcription terminator/polyadenylation signals include those derived from SV40, as described in Sambrook, et al., *supra*, as well as a bovine growth hormone terminator sequence. Introns, containing splice

donor and acceptor sites, may also be designed into the constructs for use with the present invention (Chapman et al., *Nuc. Acids Res.* (1991) 19:3979-3986).

Enhancer elements may also be used herein to increase expression levels of the mammalian constructs. Examples include the SV40 early gene enhancer, as
 5 described in Dijkema et al., *EMBO J.* (1985) 4:761, the enhancer/promoter derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus, as described in Gorman et al., *Proc. Natl. Acad. Sci. USA* (1982b) 79:6777 and elements derived from human CMV, as described in Boshart et al., *Cell* (1985) 41:521, such as elements included in the CMV intron A sequence (Chapman et al., *Nuc. Acids Res.*
 10 (1991) 19:3979-3986).

The desired synthetic Gag or Env polypeptide encoding sequences can be cloned into any number of commercially available vectors to generate expression of the polypeptide in an appropriate host system. These systems include, but are not limited to, the following: baculovirus expression {Reilly, P.R., et al., BACULOVIRUS
 15 EXPRESSION VECTORS: A LABORATORY MANUAL (1992); Beames, et al., *Biotechniques* 11:378 (1991); Pharmingen; Clontech, Palo Alto, CA)}, vaccinia expression {Earl, P. L., et al., "Expression of proteins in mammalian cells using vaccinia" In *Current Protocols in Molecular Biology* (F. M. Ausubel, et al. Eds.), Greene Publishing Associates & Wiley Interscience, New York (1991); Moss, B., et
 20 al., U.S. Patent Number 5,135,855, issued 4 August 1992}, expression in bacteria {Ausubel, F.M., et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley and Sons, Inc., Media PA; Clontech}, expression in yeast {Rosenberg, S. and Tekamp-Olson, P., U.S. Patent No. RE35,749, issued, March 17, 1998; Shuster, J.R., U.S. Patent No. 5,629,203, issued May 13, 1997; Gellissen, G., et al., *Antonie Van*
 25 *Leeuwenhoek*, 62(1-2):79-93 (1992); Romanos, M.A., et al., *Yeast* 8(6):423-488 (1992); Goeddel, D.V., *Methods in Enzymology* 185 (1990); Guthrie, C., and G.R. Fink, *Methods in Enzymology* 194 (1991)}, expression in mammalian cells {Clontech; Gibco-BRL, Ground Island, NY; e.g., Chinese hamster ovary (CHO) cell lines (Haynes, J., et al., *Nuc. Acid. Res.* 11:687-706 (1983); 1983, Lau, Y.F., et al., *Mol.*
 30 *Cell. Biol.* 4:1469-1475 (1984); Kaufman, R. J., "Selection and coamplification of heterologous genes in mammalian cells," in *Methods in Enzymology*, vol. 185, pp537-566. Academic Press, Inc., San Diego CA (1991)}, and expression in plant cells

{plant cloning vectors, Clontech Laboratories, Inc., Palo Alto, CA, and Pharmacia LKB Biotechnology, Inc., Piscataway, NJ; Hood, E., *et al.*, *J. Bacteriol.* 168:1291-1301 (1986); Nagel, R., *et al.*, *FEMS Microbiol. Lett.* 67:325 (1990); An, *et al.*, "Binary Vectors", and others in *Plant Molecular Biology Manual* A3:1-19 (1988);

5 Miki, B.L.A., *et al.*, pp.249-265, and others in *Plant DNA Infectious Agents* (Hohn, T., *et al.*, eds.) Springer-Verlag, Wien, Austria, (1987); *Plant Molecular Biology: Essential Techniques*, P.G. Jones and J.M. Sutton, New York, J. Wiley, 1997; Miglani, Gurbachan *Dictionary of Plant Genetics and Molecular Biology*, New York, Food Products Press, 1998; Henry, R. J., *Practical Applications of Plant Molecular*

10 *Biology*, New York, Chapman & Hall, 1997}.

Also included in the invention is an expression vector, containing coding sequences and expression control elements which allow expression of the coding regions in a suitable host. The control elements generally include a promoter, translation initiation codon, and translation and transcription termination sequences,

15 and an insertion site for introducing the insert into the vector. Translational control elements have been reviewed by M. Kozak (e.g., Kozak, M., *Mamm. Genome* 7(8):563-574, 1996; Kozak, M., *Biochimie* 76(9):815-821, 1994; Kozak, M., *J Cell Biol* 108(2):229-241, 1989; Kozak, M., and Shatkin, A.J., *Methods Enzymol* 60:360-375, 1979).

20 Expression in yeast systems has the advantage of commercial production. Recombinant protein production by vaccinia and CHO cell line have the advantage of being mammalian expression systems. Further, vaccinia virus expression has several advantages including the following: (i) its wide host range; (ii) faithful post-transcriptional modification, processing, folding, transport, secretion, and assembly of

25 recombinant proteins; (iii) high level expression of relatively soluble recombinant proteins; and (iv) a large capacity to accommodate foreign DNA.

The recombinantly expressed polypeptides from synthetic Gag- and/or Env-encoding expression cassettes are typically isolated from lysed cells or culture media. Purification can be carried out by methods known in the art including salt

30 fractionation, ion exchange chromatography, gel filtration, size-exclusion chromatography, size-fractionation, and affinity chromatography. Immunoaffinity

chromatography can be employed using antibodies generated based on, for example, Gag or Env antigens.

Advantages of expressing the Gag- and/or Env-containing proteins of the present invention using mammalian cells include, but are not limited to, the following: well-established protocols for scale-up production; the ability to produce VLPs; cell lines are suitable to meet good manufacturing process (GMP) standards; culture conditions for mammalian cells are known in the art.

Various forms of the different embodiments of the invention, described herein, may be combined.

10

2.2 PRODUCTION OF VIRUS-LIKE PARTICLES AND USE OF THE CONSTRUCTS OF THE PRESENT INVENTION TO CREATE PACKAGING CELL LINES.

The group-specific antigens (Gag) of human immunodeficiency virus type-1 (HIV-1) self-assemble into noninfectious virus-like particles (VLP) that are released from various eucaryotic cells by budding (reviewed by Freed, E.O., *Virology* 251:1-15, 1998). The synthetic expression cassettes of the present invention provide efficient means for the production of HIV-Gag virus-like particles (VLPs) using a variety of different cell types, including, but not limited to, mammalian cells.

Viral particles can be used as a matrix for the proper presentation of an antigen entrapped or associated therewith to the immune system of the host.

2.2.1 VLP PRODUCTION USING THE SYNTHETIC EXPRESSION CASSETTES OF THE PRESENT INVENTION

Experiments can be performed in support of the present invention to demonstrate that the synthetic expression cassettes of the present invention provide superior production of both Gag proteins and VLPs, relative to native Gag coding sequences. Further, electron microscopic evaluation of VLP production can show that free and budding immature virus particles of the expected size are produced by cells containing the synthetic expression cassettes.

Using the synthetic expression cassettes of the present invention, rather than native Gag coding sequences, for the production of virus-like particles provide several

advantages. First, VLPs can be produced in enhanced quantity making isolation and purification of the VLPs easier. Second, VLPs can be produced in a variety of cell types using the synthetic expression cassettes, in particular, mammalian cell lines can be used for VLP production, for example, CHO cells. Production using CHO cells provides (i) VLP formation; (ii) correct myristylation and budding; (iii) absence of non-mamallian cell contaminants (e.g., insect viruses and/or cells); and (iv) ease of purification. The synthetic expression cassettes of the present invention are also useful for enhanced expression in cell-types other than mammalian cell lines. For example, infection of insect cells with baculovirus vectors encoding the synthetic expression cassettes results in higher levels of total Gag protein yield and higher levels of VLP production (relative to wild-type coding sequences). Further, the final product from insect cells infected with the baculovirus-Gag synthetic expression cassettes consistently contains lower amounts of contaminating insect proteins than the final product when wild-type coding sequences are used.

VLPs can spontaneously form when the particle-forming polypeptide of interest is recombinantly expressed in an appropriate host cell. Thus, the VLPs produced using the synthetic expression cassettes of the present invention are conveniently prepared using recombinant techniques. As discussed below, the Gag polypeptide encoding synthetic expression cassettes of the present invention can include other polypeptide coding sequences of interest (for example, HIV protease, HIV polymerase, HCV core; Env; synthetic Env; see, Example 1). Expression of such synthetic expression cassettes yields VLPs comprising the Gag polypeptide, as well as, the polypeptide of interest.

Once coding sequences for the desired particle-forming polypeptides have been isolated or synthesized, they can be cloned into any suitable vector or replicon for expression. Numerous cloning vectors are known to those of skill in the art, and the selection of an appropriate cloning vector is a matter of choice. See, generally, Sambrook et al, *supra*. The vector is then used to transform an appropriate host cell. Suitable recombinant expression systems include, but are not limited to, bacterial, mammalian, baculovirus/insect, vaccinia, Semliki Forest virus (SFV), Alphaviruses (such as, Sindbis, Venezuelan Equine Encephalitis (VEE)), mammalian, yeast and

Xenopus expression systems, well known in the art. Particularly preferred expression systems are mammalian cell lines, vaccinia, Sindbis, insect and yeast systems.

For example, a number of mammalian cell lines are known in the art and include immortalized cell lines available from the American Type Culture Collection (A.T.C.C.), such as, but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), as well as others. Similarly, bacterial hosts such as *E. coli*, *Bacillus subtilis*, and *Streptococcus* spp., will find use with the present expression constructs. Yeast hosts useful in the present invention include *inter alia*, *Saccharomyces cerevisiae*, *Candida albicans*, *Candida maltosa*, *Hansenula polymorpha*, *Kluyveromyces fragilis*, *Kluyveromyces lactis*, *Pichia guillermondii*, *Pichia pastoris*, *Schizosaccharomyces pombe* and *Yarrowia lipolytica*. Insect cells for use with baculovirus expression vectors include, *inter alia*, *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni*. See, e.g., Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987).

Viral vectors can be used for the production of particles in eucaryotic cells, such as those derived from the pox family of viruses, including vaccinia virus and avian poxvirus. Additionally, a vaccinia based infection/transfection system, as described in Tomei et al., *J. Virol.* (1993) 67:4017-4026 and Selby et al., *J. Gen. Virol.* (1993) 74:1103-1113, will also find use with the present invention. In this system, cells are first infected *in vitro* with a vaccinia virus recombinant that encodes the bacteriophage T7 RNA polymerase. This polymerase displays exquisite specificity in that it only transcribes templates bearing T7 promoters. Following infection, cells are transfected with the DNA of interest, driven by a T7 promoter. The polymerase expressed in the cytoplasm from the vaccinia virus recombinant transcribes the transfected DNA into RNA which is then translated into protein by the host translational machinery. Alternately, T7 can be added as a purified protein or enzyme as in the "Progenitor" system (Studier and Moffatt, *J. Mol. Biol.* (1986) 189:113-130). The method provides for high level, transient, cytoplasmic production of large quantities of RNA and its translation product(s).

Depending on the expression system and host selected, the VLPS are produced by growing host cells transformed by an expression vector under conditions whereby

the particle-forming polypeptide is expressed and VLPs can be formed. The selection of the appropriate growth conditions is within the skill of the art. If the VLPs are formed intracellularly, the cells are then disrupted, using chemical, physical or mechanical means, which lyse the cells yet keep the VLPs substantially intact. Such methods are known to those of skill in the art and are described in, e.g., *Protein Purification Applications: A Practical Approach*, (E.L.V. Harris and S. Angal, Eds., 1990).

The particles are then isolated (or substantially purified) using methods that preserve the integrity thereof, such as, by gradient centrifugation, e.g., cesium chloride (CsCl) sucrose gradients, pelleting and the like (see, e.g., Kimbauer et al. *J. Virol.* (1993) 67:6929-6936), as well as standard purification techniques including, e.g., ion exchange and gel filtration chromatography.

VLPs produced by cells containing the synthetic expression cassettes of the present invention can be used to elicit an immune response when administered to a subject. One advantage of the present invention is that VLPs can be produced by mammalian cells carrying the synthetic expression cassettes at levels previously not possible. As discussed above, the VLPs can comprise a variety of antigens in addition to the Gag polypeptide (e.g., Gag-protease, Gag-polymerase, Env, synthetic Env, etc.). Purified VLPs, produced using the synthetic expression cassettes of the present invention, can be administered to a vertebrate subject, usually in the form of vaccine compositions. Combination vaccines may also be used, where such vaccines contain, for example, an adjuvant subunit protein (e.g., Env). Administration can take place using the VLPs formulated alone or formulated with other antigens. Further, the VLPs can be administered prior to, concurrent with, or subsequent to, delivery of the synthetic expression cassettes for DNA immunization (see below) and/or delivery of other vaccines. Also, the site of VLP administration may be the same or different as other vaccine compositions that are being administered. Gene delivery can be accomplished by a number of methods including, but are not limited to, immunization with DNA, alphavirus vectors, pox virus vectors, and vaccinia virus vectors.

VLP immune-stimulating (or vaccine) compositions can include various excipients, adjuvants, carriers, auxiliary substances, modulating agents, and the like. The immune stimulating compositions will include an amount of the VLP/antigen

sufficient to mount an immunological response. An appropriate effective amount can be determined by one of skill in the art. Such an amount will fall in a relatively broad range that can be determined through routine trials and will generally be an amount on the order of about 0.1 μ g to about 1000 μ g, more preferably about 1 μ g to about 300 μ g, of VLP/antigen.

A carrier is optionally present which is a molecule that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Examples of particulate carriers include those derived from polymethyl methacrylate polymers, as well as microparticles derived from poly(lactides) and poly(lactide-co-glycolides), known as PLG. See, e.g., Jeffery et al., *Pharm. Res.* (1993) 10:362-368; McGee JP, et al., *J Microencapsul.* 14(2):197-210, 1997; O'Hagan DT, et al., *Vaccine* 11(2):149-54, 1993. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen may be conjugated to a bacterial toxoid, such as toxoid from diphtheria, tetanus, cholera, etc., as well as toxins derived from *E. coli*.

Adjuvants may also be used to enhance the effectiveness of the compositions. Such adjuvants include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc.; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59 (International Publication No. WO 90/14837), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) RibiTM adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene,

0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particle
5 generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (IL-1, IL-2, etc.), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc.; (6) oligonucleotides or polymeric molecules encoding immunostimulatory CpG motifs (Davis, H.L., et al., *J. Immunology*
10 160:870-876, 1998; Sato, Y. et al., *Science* 273:352-354, 1996) or complexes of antigens/oligonucleotides {Polymeric molecules include double and single stranded RNA and DNA, and backbone modifications thereof, for example, methylphosphonate linkages; or (7) detoxified mutants of a bacterial ADP-
15 ribosylating toxin such as a cholera toxin (CT), a pertussis toxin (PT), or an *E. coli* heat-labile toxin (LT), particularly LT-K63 (where lysine is substituted for the wild-type amino acid at position 63) LT-R72 (where arginine is substituted for the wild-type amino acid at position 72), CT-S109 (where serine is substituted for the wild-type amino acid at position 109), and PT-K9/G129 (where lysine is substituted for the wild-type amino acid at position 9 and glycine substituted at position 129) (see, e.g.,
20 International Publication Nos. W093/13202 and W092/19265); and (8) other substances that act as immunostimulating agents to enhance the effectiveness of the composition. Further, such polymeric molecules include alternative polymer backbone structures such as, but not limited to, polyvinyl backbones (Pitha, *Biochem Biophys Acta*, 204:39, 1970a; Pitha, *Biopolymers*, 9:965, 1970b), and morpholino
25 backbones (Summerton, J., et al., U.S. Patent No. 5,142,047, issued 08/25/92; Summerton, J., et al., U.S. Patent No. 5,185,444 issued 02/09/93). A variety of other charged and uncharged polynucleotide analogs have been reported. Numerous backbone modifications are known in the art, including, but not limited to, uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoamidates, and
30 carbamates) and charged linkages (e.g., phosphorothioates and phosphorodithioates).}; and (7) other substances that act as immunostimulating agents

to enhance the effectiveness of the VLP immune-stimulating (or vaccine) composition. Alum, CpG oligonucleotides, and MF59 are preferred.

Muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamate (nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-*sn*-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), etc.

Dosage treatment with the VLP composition may be a single dose schedule or a multiple dose schedule. A multiple dose schedule is one in which a primary course of vaccination may be with 1-10 separate doses, followed by other doses given at subsequent time intervals, chosen to maintain and/or reinforce the immune response, for example at 1-4 months for a second dose, and if needed, a subsequent dose(s) after several months. The dosage regimen will also, at least in part, be determined by the need of the subject and be dependent on the judgment of the practitioner.

If prevention of disease is desired, the antigen carrying VLPs are generally administered prior to primary infection with the pathogen of interest. If treatment is desired, e.g., the reduction of symptoms or recurrences, the VLP compositions are generally administered subsequent to primary infection.

2.2.2 USING THE SYNTHETIC EXPRESSION CASSETTES OF THE PRESENT INVENTION TO CREATE PACKAGING CELL LINES

A number of viral based systems have been developed for use as gene transfer vectors for mammalian host cells. For example, retroviruses (in particular, lentiviral vectors) provide a convenient platform for gene delivery systems. A coding sequence of interest (for example, a sequence useful for gene therapy applications) can be inserted into a gene delivery vector and packaged in retroviral particles using techniques known in the art. Recombinant virus can then be isolated and delivered to cells of the subject either *in vivo* or *ex vivo*. A number of retroviral systems have been described, including, for example, the following: (U.S. Patent No. 5,219,740; Miller et al. (1989) *BioTechniques* 7:980; Miller, A.D. (1990) *Human Gene Therapy* 1:5; Scarpa et al. (1991) *Virology* 180:849; Burns et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:8033; Boris-Lawrie et al. (1993) *Cur. Opin. Genet. Develop.* 3:102; GB 2200651; EP 0415731; EP 0345242; WO 89/02468; WO 89/05349; WO 89/09271; WO

90/02806; WO 90/07936; WO 90/07936; WO 94/03622; WO 93/25698; WO
 93/25234; WO 93/11230; WO 93/10218; WO 91/02805; in U.S. 5,219,740; U.S.
 4,405,712; U.S. 4,861,719; U.S. 4,980,289 and U.S. 4,777,127; in U.S. Serial No.
 07/800,921; and in Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res*
 5 53:962-967; Ram (1993) *Cancer Res* 53:83-88; Takamiya (1992) *J Neurosci Res*
33:493-503; Baba (1993) *J Neurosurg* 79:729-735; Mann (1983) *Cell* 33:153; Cane
 (1984) *Proc Natl Acad Sci USA* 81:6349; and Miller (1990) *Human Gene Therapy* 1.

In other embodiments, gene transfer vectors can be constructed to encode a
 cytokine or other immunomodulatory molecule. For example, nucleic acid sequences
 10 encoding native IL-2 and gamma-interferon can be obtained as described in US Patent
 Nos. 4,738,927 and 5,326,859, respectively, while useful muteins of these proteins
 can be obtained as described in U.S. Patent No. 4,853,332. Nucleic acid sequences
 encoding the short and long forms of mCSF can be obtained as described in US Patent
 Nos. 4,847,201 and 4,879,227, respectively. In particular aspects of the invention,
 15 retroviral vectors expressing cytokine or immunomodulatory genes can be produced
 as described herein (for example, employing the packaging cell lines of the present
 invention) and in International Application No. PCT US 94/02951, entitled
 "Compositions and Methods for Cancer Immunotherapy."

Examples of suitable immunomodulatory molecules for use herein include the
 20 following: IL-1 and IL-2 (Karupiah et al. (1990) *J. Immunology* 144:290-298, Weber
 et al. (1987) *J. Exp. Med.* 166:1716-1733, Gansbacher et al. (1990) *J. Exp. Med.*
172:1217-1224, and U.S. Patent No. 4,738,927); IL-3 and IL-4 (Tepper et al. (1989)
Cell 57:503-512, Golumbek et al. (1991) *Science* 254:713-716, and U.S. Patent No.
 5,017,691); IL-5 and IL-6 (Brakenhof et al. (1987) *J. Immunol.* 139:4116-4121, and
 25 International Publication No. WO 90/06370); IL-7 (U.S. Patent No. 4,965,195); IL-8,
 IL-9, IL-10, IL-11, IL-12, and IL-13 (*Cytokine Bulletin*, Summer 1994); IL-14 and
 IL-15; alpha interferon (Finter et al. (1991) *Drugs* 42:749-765, U.S. Patent Nos.
 4,892,743 and 4,966,843, International Publication No. WO 85/02862, Nagata et al.
 (1980) *Nature* 284:316-320, Familletti et al. (1981) *Methods in Enz.* 78:387-394, Twu
 30 et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:2046-2050, and Faktor et al. (1990)
Oncogene 5:867-872); beta-interferon (Seif et al. (1991) *J. Virol.* 65:664-671);
 gamma-interferons (Radford et al. (1991) *The American Society of Hepatology*

20082015, Watanabe et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:9456-9460,
Gansbacher et al. (1990) *Cancer Research* 50:7820-7825, Maio et al. (1989) *Can.*
Immunol. Immunother. 30:34-42, and U.S. Patent Nos. 4,762,791 and 4,727,138); G-
CSF (U.S. Patent Nos. 4,999,291 and 4,810,643); GM-CSF (International Publication
5 No. WO 85/04188).

Immunomodulatory factors may also be agonists, antagonists, or ligands for
these molecules. For example, soluble forms of receptors can often behave as
antagonists for these types of factors, as can mutated forms of the factors themselves.

Nucleic acid molecules that encode the above-described substances, as well as
10 other nucleic acid molecules that are advantageous for use within the present
invention, may be readily obtained from a variety of sources, including, for example,
depositories such as the American Type Culture Collection, or from commercial
sources such as British Bio-Technology Limited (Cowley, Oxford England).
Representative examples include BBG 12 (containing the GM-CSF gene coding for
15 the mature protein of 127 amino acids), BBG 6 (which contains sequences encoding
gamma interferon), A.T.C.C. Deposit No. 39656 (which contains sequences encoding
TNF), A.T.C.C. Deposit No. 20663 (which contains sequences encoding alpha-
interferon), A.T.C.C. Deposit Nos. 31902, 31902 and 39517 (which contain
sequences encoding beta-interferon), A.T.C.C. Deposit No. 67024 (which contains a
20 sequence which encodes Interleukin-1b), A.T.C.C. Deposit Nos. 39405, 39452,
39516, 39626 and 39673 (which contain sequences encoding Interleukin-2), A.T.C.C.
Deposit Nos. 59399, 59398, and 67326 (which contain sequences encoding
Interleukin-3), A.T.C.C. Deposit No. 57592 (which contains sequences encoding
Interleukin-4), A.T.C.C. Deposit Nos. 59394 and 59395 (which contain sequences
25 encoding Interleukin-5), and A.T.C.C. Deposit No. 67153 (which contains sequences
encoding Interleukin-6).

Plasmids containing cytokine genes or immunomodulatory genes
(International Publication Nos. WO 94/02951 and WO) can be digested with
appropriate restriction enzymes, and DNA fragments containing the particular gene of
30 interest can be inserted into a gene transfer vector using standard molecular biology
techniques. (See, e.g., Sambrook et al., *supra.*, or Ausbel et al. (eds) *Current*
Protocols in Molecular Biology, Greene Publishing and Wiley-Interscience).

Polynucleotide sequences coding for the above-described molecules can be obtained using recombinant methods, such as by screening cDNA and genomic libraries from cells expressing the gene, or by deriving the gene from a vector known to include the same. For example, plasmids which contain sequences that encode
5 altered cellular products may be obtained from a depository such as the A.T.C.C., or from commercial sources. Plasmids containing the nucleotide sequences of interest can be digested with appropriate restriction enzymes, and DNA fragments containing the nucleotide sequences can be inserted into a gene transfer vector using standard molecular biology techniques.

10 Alternatively, cDNA sequences for use with the present invention may be obtained from cells which express or contain the sequences, using standard techniques, such as phenol extraction and PCR of cDNA or genomic DNA. See, e.g., Sambrook et al., *supra*, for a description of techniques used to obtain and isolate DNA. Briefly, mRNA from a cell which expresses the gene of interest can be reverse
15 transcribed with reverse transcriptase using oligo-dT or random primers. The single stranded cDNA may then be amplified by PCR (see U.S. Patent Nos. 4,683,202, 4,683,195 and 4,800,159, see also *PCR Technology: Principles and Applications for DNA Amplification*, Erlich (ed.), Stockton Press, 1989)) using oligonucleotide primers complementary to sequences on either side of desired sequences.

20 The nucleotide sequence of interest can also be produced synthetically, rather than cloned, using a DNA synthesizer (e.g., an Applied Biosystems Model 392 DNA Synthesizer, available from ABI, Foster City, California). The nucleotide sequence can be designed with the appropriate codons for the expression product desired. The complete sequence is assembled from overlapping oligonucleotides prepared by
25 standard methods and assembled into a complete coding sequence. See, e.g., Edge (1981) *Nature* 292:756; Nambair et al. (1984) *Science* 223:1299; Jay et al. (1984) *J. Biol. Chem.* 259:6311.

The synthetic expression cassettes of the present invention can be employed in the construction of packaging cell lines for use with retroviral vectors.

30 One type of retrovirus, the murine leukemia virus, or "MLV", has been widely utilized for gene therapy applications (see generally Mann et al. (*Cell* 33:153, 1993),

Cane and Mulligan (*Proc. Nat'l. Acad. Sci. USA* 81:6349, 1984), and Miller et al., *Human Gene Therapy* 1:5-14, 1990.

Lentiviral vectors typically, comprise a 5' lentiviral LTR, a tRNA binding site, a packaging signal, a promoter operably linked to one or more genes of interest, an
5 origin of second strand DNA synthesis and a 3' lentiviral LTR, wherein the lentiviral vector contains a nuclear transport element. The nuclear transport element may be located either upstream (5') or downstream (3') of a coding sequence of interest (for example, a synthetic Gag or Env expression cassette of the present invention). Within certain embodiments, the nuclear transport element is not RRE. Within one
10 embodiment the packaging signal is an extended packaging signal. Within other embodiments the promoter is a tissue specific promoter, or, alternatively, a promoter such as CMV. Within other embodiments, the lentiviral vector further comprises an internal ribosome entry site.

A wide variety of lentiviruses may be utilized within the context of the present
15 invention, including for example, lentiviruses selected from the group consisting of HIV, HIV-1, HIV-2, FIV and SIV.

In one embodiment of the present invention synthetic Gag-polymerase expression cassettes are provided comprising a promoter and a sequence encoding synthetic Gag-polymerase and at least one of vpr, vpu, nef or vif, wherein the
20 promoter is operably linked to Gag-polymerase and vpr, vpu, nef or vif.

Within yet another aspect of the invention, host cells (eg., packaging cell lines) are provided which contain any of the expression cassettes described herein. For example, within one aspect packaging cell line are provided comprising an expression cassette that comprises a sequence encoding synthetic Gag-polymerase,
25 and a nuclear transport element, wherein the promoter is operably linked to the sequence encoding Gag-polymerase. Packaging cell lines may further comprise a promoter and a sequence encoding tat, rev, or an envelope, wherein the promoter is operably linked to the sequence encoding tat, rev, Env or modified Env proteins. The packaging cell line may further comprise a sequence encoding any one or more of nef,
30 vif, vpu or vpr.

In one embodiment, the expression cassette (carrying, for example, the synthetic Gag-polymerase) is stably integrated. The packaging cell line, upon

introduction of a lentiviral vector, typically produces particles. The promoter regulating expression of the synthetic expression cassette may be inducible. Typically, the packaging cell line, upon introduction of a lentiviral vector, produces particles that are essentially free of replication competent virus.

5 Packaging cell lines are provided comprising an expression cassette which directs the expression of a synthetic *Gag-polymerase* gene or comprising an expression cassette which directs the expression of a synthetic Env genes described herein. (See, also, Andre, S., et al., *Journal of Virology* 72(2):1497-1503, 1998; Haas, J., et al., *Current Biology* 6(3):315-324, 1996) for a description of other modified Env
10 sequences). A lentiviral vector is introduced into the packaging cell line to produce a vector producing cell line.

As noted above, lentiviral vectors can be designed to carry or express a selected gene(s) or sequences of interest. Lentiviral vectors may be readily constructed from a wide variety of lentiviruses (*see* RNA Tumor Viruses, Second
15 Edition, Cold Spring Harbor Laboratory, 1985). Representative examples of lentiviruses included HIV, HIV-1, HIV-2, FIV and SIV. Such lentiviruses may either be obtained from patient isolates, or, more preferably, from depositories or collections such as the American Type Culture Collection, or isolated from known sources using available techniques.

20 Portions of the lentiviral gene delivery vectors (or vehicles) may be derived from different viruses. For example, in a given recombinant lentiviral vector, LTRs may be derived from an HIV, a packaging signal from SIV, and an origin of second strand synthesis from HrV-2. Lentiviral vector constructs may comprise a 5' lentiviral LTR, a tRNA binding site, a packaging signal, one or more heterologous sequences,
25 an origin of second strand DNA synthesis and a 3' LTR, wherein said lentiviral vector contains a nuclear transport element that is not RRE.

Briefly, Long Terminal Repeats ("LTRs") are subdivided into three elements, designated U5, R and U3. These elements contain a variety of signals which are responsible for the biological activity of a retrovirus, including for example, promoter
30 and enhancer elements which are located within U3. LTRs may be readily identified in the provirus (integrated DNA form) due to their precise duplication at either end of the genome. As utilized herein, a 5' LTR should be understood to include a 5'

promoter element and sufficient LTR sequence to allow reverse transcription and integration of the DNA form of the vector. The 3' LTR should be understood to include a polyadenylation signal, and sufficient LTR sequence to allow reverse transcription and integration of the DNA form of the vector.

5 The tRNA binding site and origin of second strand DNA synthesis are also important for a retrovirus to be biologically active, and may be readily identified by one of skill in the art. For example, retroviral tRNA binds to a tRNA binding site by Watson-Crick base pairing, and is carried with the retrovirus genome into a viral particle. The tRNA is then utilized as a primer for DNA synthesis by reverse transcriptase. The tRNA binding site may be readily identified based upon its location just downstream from the 5'LTR. Similarly, the origin of second strand DNA synthesis is, as its name implies, important for the second strand DNA synthesis of a retrovirus. This region, which is also referred to as the poly-purine tract, is located just upstream of the 3'LTR.

15 In addition to a 5' and 3' LTR, tRNA binding site, and origin of second strand DNA synthesis, recombinant retroviral vector constructs may also comprise a packaging signal, as well as one or more genes or coding sequences of interest. In addition, the lentiviral vectors have a nuclear transport element which, in preferred embodiments is not RRE. Representative examples of suitable nuclear transport elements include the element in Rous sarcoma virus (Ogert, et al., *J Virol* 70, 3834-3843, 1996), the element in Rous sarcoma virus (Liu & Mertz, *Genes & Dev.*, 9, 1766-1789, 1995) and the element in the genome of simian retrovirus type I (Zolotukhin, et al., *J Virol.* 68, 7944-7952, 1994). Other potential elements include the elements in the histone gene (Kedes, *Annu. Rev. Biochem.* 48, 837-870, 1970), the α -interferon gene (Nagata et al., *Nature* 287, 401-408, 1980), the β -adrenergic receptor gene (Koilkka, et al., *Nature* 329, 75-79, 1987), and the c-Jun gene (Hattorie, et al., *Proc. Natl. Acad. Sci. USA* 85, 9148-9152, 1988).

25 Recombinant lentiviral vector constructs typically lack both *Gag-polymerase* and *Env* coding sequences. Recombinant lentiviral vector typically contain less than 20, preferably 15, more preferably 10, and most preferably 8 consecutive nucleotides found in *Gag-polymerase* and *Env* genes. One advantage of the present invention is that the synthetic *Gag-polymerase* expression cassettes, which can be used to

construct packaging cell lines for the recombinant retroviral vector constructs, have little homology to wild-type Gag-polymerase sequences and thus considerably reduce or eliminate the possibility of homologous recombination between the synthetic and wild-type sequences.

5 Lentiviral vectors may also include tissue-specific promoters to drive expression of one or more genes or sequences of interest.

 Lentiviral vector constructs may be generated such that more than one gene of interest is expressed. This may be accomplished through the use of di- or oligo-cistronic cassettes (e.g., where the coding regions are separated by 80 nucleotides or less, *see generally* Levin et al., *Gene* 108:167-174, 1991), or through the use of
10 Internal Ribosome Entry Sites ("IRES").

 Packaging cell lines suitable for use with the above described recombinant retroviral vector constructs may be readily prepared given the disclosure provided herein. Briefly, the parent cell line from which the packaging cell line is derived can
15 be selected from a variety of mammalian cell lines, including for example, 293, RD, COS-7, CHO, BHK, VERO, HT1080, and myeloma cells.

 After selection of a suitable host cell for the generation of a packaging cell line, one or more expression cassettes are introduced into the cell line in order to complement or supply in *trans* components of the vector which have been deleted.

20 Representative examples of suitable expression cassettes have been described herein and include synthetic Env, synthetic Gag, synthetic Gag-protease, and synthetic Gag-polymerase expression cassettes, which comprise a promoter and a sequence encoding, e.g., Gag-polymerase and at least one of vpr, vpu, nef or vif, wherein the promoter is operably linked to Gag-polymerase and vpr, vpu, nef or vif. As described
25 above, the native and/or modified Env coding sequences may also be utilized in these expression cassettes.

 Utilizing the above-described expression cassettes, a wide variety of packaging cell lines can be generated. For example, within one aspect packaging cell line are provided comprising an expression cassette that comprises a sequence
30 encoding synthetic Gag-polymerase, and a nuclear transport element, wherein the promoter is operably linked to the sequence encoding Gag-polymerase. Within other aspects, packaging cell lines are provided comprising a promoter and a sequence

encoding tat, rev, Env, or other HIV antigens or epitopes derived therefrom, wherein the promoter is operably linked to the sequence encoding tat, rev, Env, or the HIV antigen or epitope. Within further embodiments, the packaging cell line may comprise a sequence encoding any one or more of nef, vif, vpu or vpr. For example, 5 the packaging cell line may contain only nef, vif, vpu, or vpr alone, nef and vif, nef and vpu, nef and vpr, vif and vpu, vif and vpr, vpu and vpr, nef vif and vpu, nef vif and vpr, nef vpu and vpr, vif vpu and vpr, or, all four of nef vif vpu and vpr.

In one embodiment, the expression cassette is stably integrated. Within another embodiment, the packaging cell line, upon introduction of a lentiviral vector, 10 produces particles. Within further embodiments the promoter is inducible. Within certain preferred embodiments of the invention, the packaging cell line, upon introduction of a lentiviral vector, produces particles that are free of replication competent virus.

The synthetic cassettes containing optimized coding sequences are transfected 15 into a selected cell line. Transfected cells are selected that (i) carry, typically, integrated, stable copies of the Gag, Pol, and Env coding sequences, and (ii) are expressing acceptable levels of these polypeptides (expression can be evaluated by methods known in the prior art, e.g., see Examples 1-4). The ability of the cell line to produce VLPs may also be verified.

20 A sequence of interest is constructed into a suitable viral vector as discussed above. This defective virus is then transfected into the packaging cell line. The packaging cell line provides the viral functions necessary for producing virus-like particles into which the defective viral genome, containing the sequence of interest, are packaged. These VLPs are then isolated and can be used, for example, in gene 25 delivery or gene therapy.

Further, such packaging cell lines can also be used to produce VLPs alone, which can, for example, be used as adjuvants for administration with other antigens or in vaccine compositions. Also, co-expression of a selected sequence of interest encoding a polypeptide (for example, an antigen) in the packaging cell line can also 30 result in the entrapment and/or association of the selected polypeptide in/with the VLPs.

2.3 DNA IMMUNIZATION AND GENE DELIVERY

A variety of HIV polypeptide antigens, particularly Type C HIV antigens, can be used in the practice of the present invention. HIV antigens can be included in DNA immunization constructs containing, for example, a synthetic Gag expression cassette fused in-frame to a coding sequence for the polypeptide antigen, where
5 expression of the construct results in VLPs presenting the antigen of interest.

HIV antigens of particular interest to be used in the practice of the present invention include tat, rev, nef, vif, vpu, vpr, and other HIV antigens or epitopes derived therefrom. For example, the packaging cell line may contain only nef, and
10 HIV-1 (also known as HTLV-III, LAV, ARV, etc.), including, but not limited to, antigens such as gp120, gp41, gp160 (both native and modified); Gag; and pol from a variety of isolates including, but not limited to, HIV_{IIIB}, HIV_{SF2}, HIV-1_{SF162}, HIV-1_{SF170}, HIV_{LAV}, HIV_{LAI}, HIV_{MN}, HIV-1_{CM235}, HIV-1_{US4}, other HIV-1 strains from diverse subtypes (e.g., subtypes, A through G, and O), HIV-2 strains and diverse subtypes
15 (e.g., HIV-2_{UC1} and HIV-2_{UC2}). See, e.g., Myers, et al., Los Alamos Database, Los Alamos National Laboratory, Los Alamos, New Mexico; Myers, et al., *Human Retroviruses and Aids*, 1990, Los Alamos, New Mexico: Los Alamos National Laboratory.

To evaluate efficacy, DNA immunization using synthetic expression cassettes
20 of the present invention can be performed, for instance as described in Example 4. Mice are immunized with both the Gag (and/or Env) synthetic expression cassette and the Gag (and/or Env) wild type expression cassette. Mouse immunizations with plasmid-DNAs will show that the synthetic expression cassettes provide a clear improvement of immunogenicity relative to the native expression cassettes. Also, the
25 second boost immunization will induce a secondary immune response, for example, after approximately two weeks. Further, the results of CTL assays will show increased potency of synthetic Gag (and/or Env) expression cassettes for induction of cytotoxic T-lymphocyte (CTL) responses by DNA immunization.

It is readily apparent that the subject invention can be used to mount an
30 immune response to a wide variety of antigens and hence to treat or prevent a HIV infection, particularly Type C HIV infection.

2.3.1 DELIVERY OF THE SYNTHETIC EXPRESSION CASSETTES OF THE PRESENT INVENTION

Polynucleotide sequences coding for the above-described molecules can be obtained using recombinant methods, such as by screening cDNA and genomic
5 libraries from cells expressing the gene, or by deriving the gene from a vector known to include the same. Furthermore, the desired gene can be isolated directly from cells and tissues containing the same, using standard techniques, such as phenol extraction and PCR of cDNA or genomic DNA. See, e.g., Sambrook et al., *supra*, for a description of techniques used to obtain and isolate DNA. The gene of interest can
10 also be produced synthetically, rather than cloned. The nucleotide sequence can be designed with the appropriate codons for the particular amino acid sequence desired. In general, one will select preferred codons for the intended host in which the sequence will be expressed. The complete sequence is assembled from overlapping oligonucleotides prepared by standard methods and assembled into a complete coding
15 sequence. See, e.g., Edge, *Nature* (1981) 292:756; Nambair et al., *Science* (1984) 223:1299; Jay et al., *J. Biol. Chem.* (1984) 259:6311; Stemmer, W.P.C., (1995) *Gene* 164:49-53.

Next, the gene sequence encoding the desired antigen can be inserted into a vector containing a synthetic Gag or synthetic Env expression cassette of the present
20 invention. The antigen is inserted into the synthetic Gag coding sequence such that when the combined sequence is expressed it results in the production of VLPs comprising the Gag polypeptide and the antigen of interest, e.g., Env (native or modified) or other antigen derived from HIV. Insertions can be made within the coding sequence or at either end of the coding sequence (5', amino terminus of the
25 expressed Gag polypeptide; or 3', carboxy terminus of the expressed Gag polypeptide)(Wagner, R., et al., *Arch Virol.* 127:117-137, 1992; Wagner, R., et al., *Virology* 200:162-175, 1994; Wu, X., et al., *J. Virol.* 69(6):3389-3398, 1995; Wang, C-T., et al., *Virology* 200:524-534, 1994; Chazal, N., et al., *Virology* 68(1):111-122, 1994; Griffiths, J.C., et al., *J. Virol.* 67(6):3191-3198, 1993; Reicin, A.S., et al., *J.*
30 *Virol.* 69(2):642-650, 1995).

Up to 50% of the coding sequences of p55Gag can be deleted without affecting the assembly to virus-like particles and expression efficiency (Borsetti, A.,

et al, *J. Virol.* 72(11):9313-9317, 1998; Gamier, L., et al., *J Virol* 72(6):4667-4677, 1998; Zhang, Y., et al., *J Virol* 72(3):1782-1789, 1998; Wang, C., et al., *J Virol* 72(10): 7950-7959, 1998). In one embodiment of the present invention, immunogenicity of the high level expressing synthetic Gag expression cassettes can be increased by the insertion of different structural or non-structural HIV antigens, multiepitope cassettes, or cytokine sequences into deleted regions of Gag sequence. Such deletions may be generated following the teachings of the present invention and information available to one of ordinary skill in the art. One possible advantage of this approach, relative to using full-length sequences fused to heterologous polypeptides, can be higher expression/secretion efficiency of the expression product.

When sequences are added to the amino terminal end of Gag, the polynucleotide can contain coding sequences at the 5' end that encode a signal for addition of a myristic moiety to the Gag-containing polypeptide (e.g., sequences that encode Met-Gly).

The ability of Gag-containing polypeptide constructs to form VLPs can be empirically determined following the teachings of the present specification.

Gag/antigen (e.g., Gag/Env) synthetic expression cassettes include control elements operably linked to the coding sequence, which allow for the expression of the gene *in vivo* in the subject species. For example, typical promoters for mammalian cell expression include the SV40 early promoter, a CMV promoter such as the CMV immediate early promoter, the mouse mammary tumor virus LTR promoter, the adenovirus major late promoter (Ad MLP), and the herpes simplex virus promoter, among others. Other nonviral promoters, such as a promoter derived from the murine metallothionein gene, will also find use for mammalian expression.

Typically, transcription termination and polyadenylation sequences will also be present, located 3' to the translation stop codon. Preferably, a sequence for optimization of initiation of translation, located 5' to the coding sequence, is also present. Examples of transcription terminator/polyadenylation signals include those derived from SV40, as described in Sambrook et al., *supra*, as well as a bovine growth hormone terminator sequence.

Enhancer elements may also be used herein to increase expression levels of the mammalian constructs. Examples include the SV40 early gene enhancer, as

described in Dijkema et al., *EMBO J.* (1985) 4:761, the enhancer/promoter derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus, as described in Gorman et al., *Proc. Natl. Acad. Sci. USA* (1982b) 79:6777 and elements derived from human CMV, as described in Boshart et al., *Cell* (1985) 41:521, such as
5 elements included in the CMV intron A sequence.

Furthermore, plasmids can be constructed which include a chimeric antigen-coding gene sequences, encoding, e.g., multiple antigens/epitopes of interest, for example derived from more than one viral isolate.

Typically the antigen coding sequences precede or follow the synthetic coding
10 sequence and the chimeric transcription unit will have a single open reading frame encoding both the antigen of interest and the synthetic Gag coding sequences. Alternatively, multi-cistronic cassettes (e.g., bi-cistronic cassettes) can be constructed allowing expression of multiple antigens from a single mRNA using the EMCV IRES, or the like.

15 Once complete, the constructs are used for nucleic acid immunization using standard gene delivery protocols. Methods for gene delivery are known in the art. See, e.g., U.S. Patent Nos. 5,399,346, 5,580,859, 5,589,466. Genes can be delivered either directly to the vertebrate subject or, alternatively, delivered *ex vivo*, to cells derived from the subject and the cells reimplanted in the subject.

20 A number of viral based systems have been developed for gene transfer into mammalian cells. For example, retroviruses provide a convenient platform for gene delivery systems. Selected sequences can be inserted into a vector and packaged in retroviral particles using techniques known in the art. The recombinant virus can then be isolated and delivered to cells of the subject either *in vivo* or *ex vivo*. A number of
25 retroviral systems have been described (U.S. Patent No. 5,219,740; Miller and Rosman, *BioTechniques* (1989) 7:980-990; Miller, A.D., *Human Gene Therapy* (1990) 1:5-14; Scarpa et al., *Virology* (1991) 180:849-852; Burns et al., *Proc. Natl. Acad. Sci. USA* (1993) 90:8033-8037; and Boris-Lawrie and Temin, *Cur. Opin. Genet. Develop.* (1993) 3:102-109.

30 A number of adenovirus vectors have also been described. Unlike retroviruses which integrate into the host genome, adenoviruses persist extrachromosomally thus minimizing the risks associated with insertional mutagenesis (Haj-Ahmad and

Graham, *J. Virol.* (1986) 57:267-274; Bett et al., *J. Virol.* (1993) 67:5911-5921; Mittereder et al., *Human Gene Therapy* (1994) 5:717-729; Seth et al., *J. Virol.* (1994) 68:933-940; Barr et al., *Gene Therapy* (1994) 1:51-58; Berkner, K.L. *BioTechniques* (1988) 6:616-629; and Rich et al., *Human Gene Therapy* (1993) 4:461-476).

5 Additionally, various adeno-associated virus (AAV) vector systems have been developed for gene delivery. AAV vectors can be readily constructed using techniques well known in the art. See, e.g., U.S. Patent Nos. 5,173,414 and 5,139,941; International Publication Nos. WO 92/01070 (published 23 January 1992) and WO 93/03769 (published 4 March 1993); Lebkowski et al., *Molec. Cell. Biol.* (1988) 8:3988-3996; Vincent et al., *Vaccines 90* (1990) (Cold Spring Harbor Laboratory Press); Carter, B.J. *Current Opinion in Biotechnology* (1992) 3:533-539; Muzyczka, N. *Current Topics in Microbiol. and Immunol.* (1992) 158:97-129; Kotin, R.M. *Human Gene Therapy* (1994) 5:793-801; Shelling and Smith, *Gene Therapy* (1994) 1:165-169; and Zhou et al., *J. Exp. Med.* (1994) 179:1867-1875.

15 Another vector system useful for delivering the polynucleotides of the present invention is the enterically administered recombinant poxvirus vaccines described by Small, Jr., P.A., et al. (U.S. Patent No. 5,676,950, issued October 14, 1997).

 Additional viral vectors which will find use for delivering the nucleic acid molecules encoding the antigens of interest include those derived from the pox family of viruses, including vaccinia virus and avian poxvirus. By way of example, vaccinia virus recombinants expressing the genes can be constructed as follows. The DNA encoding the particular synthetic Gag/ or Env/antigen coding sequence is first inserted into an appropriate vector so that it is adjacent to a vaccinia promoter and flanking vaccinia DNA sequences, such as the sequence encoding thymidine kinase (TK).

20 This vector is then used to transfect cells which are simultaneously infected with vaccinia. Homologous recombination serves to insert the vaccinia promoter plus the gene encoding the coding sequences of interest into the viral genome. The resulting TK recombinant can be selected by culturing the cells in the presence of 5-bromodeoxyuridine and picking viral plaques resistant thereto.

30 Alternatively, avipoxviruses, such as the fowlpox and canarypox viruses, can also be used to deliver the genes. Recombinant avipox viruses, expressing immunogens from mammalian pathogens, are known to confer protective immunity

when administered to non-avian species. The use of an avipox vector is particularly desirable in human and other mammalian species since members of the avipox genus can only productively replicate in susceptible avian species and therefore are not infective in mammalian cells. Methods for producing recombinant avipoxviruses are known in the art and employ genetic recombination, as described above with respect to the production of vaccinia viruses. See, e.g., WO 91/12882; WO 89/03429; and WO 92/03545.

Molecular conjugate vectors, such as the adenovirus chimeric vectors described in Michael et al., *J. Biol. Chem.* (1993) 268:6866-6869 and Wagner et al., *Proc. Natl. Acad. Sci. USA* (1992) 89:6099-6103, can also be used for gene delivery.

Members of the Alphavirus genus, such as, but not limited to, vectors derived from the Sindbis, Semliki Forest, and Venezuelan Equine Encephalitis viruses, will also find use as viral vectors for delivering the polynucleotides of the present invention (for example, a synthetic Gag-polypeptide encoding expression cassette). For a description of Sindbis-virus derived vectors useful for the practice of the instant methods, see, Dubensky et al., *J. Virol.* (1996) 70:508-519; and International Publication Nos. WO 95/07995 and WO 96/17072; as well as, Dubensky, Jr., T.W., et al., U.S. Patent No. 5,843,723, issued December 1, 1998, and Dubensky, Jr., T.W., U.S. Patent No. 5,789,245, issued August 4, 1998.

A vaccinia based infection/transfection system can be conveniently used to provide for inducible, transient expression of the coding sequences of interest in a host cell. In this system, cells are first infected *in vitro* with a vaccinia virus recombinant that encodes the bacteriophage T7 RNA polymerase. This polymerase displays exquisite specificity in that it only transcribes templates bearing T7 promoters. Following infection, cells are transfected with the polynucleotide of interest, driven by a T7 promoter. The polymerase expressed in the cytoplasm from the vaccinia virus recombinant transcribes the transfected DNA into RNA which is then translated into protein by the host translational machinery. The method provides for high level, transient, cytoplasmic production of large quantities of RNA and its translation products. See, e.g., Elroy-Stein and Moss, *Proc. Natl. Acad. Sci. USA* (1990) 87:6743-6747; Fuerst et al., *Proc. Natl. Acad. Sci. USA* (1986) 83:8122-8126.

As an alternative approach to infection with vaccinia or avipox virus recombinants, or to the delivery of genes using other viral vectors, an amplification system can be used that will lead to high level expression following introduction into host cells. Specifically, a T7 RNA polymerase promoter preceding the coding region for T7 RNA polymerase can be engineered. Translation of RNA derived from this template will generate T7 RNA polymerase which in turn will transcribe more template. Concomitantly, there will be a cDNA whose expression is under the control of the T7 promoter. Thus, some of the T7 RNA polymerase generated from translation of the amplification template RNA will lead to transcription of the desired gene. Because some T7 RNA polymerase is required to initiate the amplification, T7 RNA polymerase can be introduced into cells along with the template(s) to prime the transcription reaction. The polymerase can be introduced as a protein or on a plasmid encoding the RNA polymerase. For a further discussion of T7 systems and their use for transforming cells, see, e.g., International Publication No. WO 94/26911; Studier and Moffatt, *J. Mol. Biol.* (1986) 189:113-130; Deng and Wolff, *Gene* (1994) 143:245-249; Gao et al., *Biochem. Biophys. Res. Commun.* (1994) 200:1201-1206; Gao and Huang, *Nuc. Acids Res.* (1993) 21:2867-2872; Chen et al., *Nuc. Acids Res.* (1994) 22:2114-2120; and U.S. Patent No. 5,135,855.

A synthetic Gag- and/or Env-containing expression cassette of interest can also be delivered without a viral vector. For example, the synthetic expression cassette can be packaged in liposomes prior to delivery to the subject or to cells derived therefrom. Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed DNA to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight, *Biochim. Biophys. Acta.* (1991) 1097:1-17; Straubinger et al., in *Methods of Enzymology* (1983), Vol. 101, pp. 512-527.

Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations, with cationic liposomes particularly preferred. Cationic liposomes have been shown to mediate intracellular delivery of plasmid DNA (Felgner et al., *Proc. Natl. Acad. Sci.*

USA (1987) 84:7413-7416); mRNA (Malone et al., *Proc. Natl. Acad. Sci. USA* (1989) 86:6077-6081); and purified transcription factors (Debs et al., *J. Biol. Chem.* (1990) 265:10189-10192), in functional form.

Cationic liposomes are readily available. For example, N[1-2,3-dioleoyloxy)propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available
5 under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See, also, Felgner et al., *Proc. Natl. Acad. Sci. USA* (1987) 84:7413-7416). Other commercially available lipids include (DDAB/DOPE) and DOTAP/DOPE (Boehringer). Other cationic liposomes can be prepared from readily available materials using techniques
10 well known in the art. See, e.g., Szoka et al., *Proc. Natl. Acad. Sci. USA* (1978) 75:4194-4198; PCT Publication No. WO 90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

Similarly, anionic and neutral liposomes are readily available, such as, from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily
15 available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphosphatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate ratios. Methods for making liposomes using these
20 materials are well known in the art.

The liposomes can comprise multilammellar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See, e.g., Straubinger et al., in *METHODS OF IMMUNOLOGY* (1983), Vol. 101, pp.
25 512-527; Szoka et al., *Proc. Natl. Acad. Sci. USA* (1978) 75:4194-4198; Papahadjopoulos et al., *Biochim. Biophys. Acta* (1975) 394:483; Wilson et al., *Cell* (1979) 17:77; Deamer and Bangham, *Biochim. Biophys. Acta* (1976) 443:629; Ostro et al., *Biochem. Biophys. Res. Commun.* (1977) 76:836; Fraley et al., *Proc. Natl. Acad. Sci. USA* (1979) 76:3348; Enoch and Strittmatter, *Proc. Natl. Acad. Sci. USA*
30 (1979) 76:145; Fraley et al., *J. Biol. Chem.* (1980) 255:10431; Szoka and Papahadjopoulos, *Proc. Natl. Acad. Sci. USA* (1978) 75:145; and Schaefer-Ridder et al., *Science* (1982) 215:166.

The DNA and/or protein antigen(s) can also be delivered in cochleate lipid compositions similar to those described by Papahadjopoulos et al., *Biochem. Biophys. Acta.* (1975) 394:483-491. See, also, U.S. Patent Nos. 4,663,161 and 4,871,488.

The synthetic expression cassette of interest may also be encapsulated, adsorbed to, or associated with, particulate carriers. Such carriers present multiple copies of a selected antigen to the immune system and promote trapping and retention of antigens in local lymph nodes. The particles can be phagocytosed by macrophages and can enhance antigen presentation through cytokine release. Examples of particulate carriers include those derived from polymethyl methacrylate polymers, as well as microparticles derived from poly(lactides) and poly(lactide-co-glycolides), known as PLG. See, e.g., Jeffery et al., *Pharm. Res.* (1993) 10:362-368; McGee JP, et al., *J Microencapsul.* 14(2):197-210, 1997; O'Hagan DT, et al., *Vaccine* 11(2):149-54, 1993. Suitable microparticles may also be manufactured in the presence of charged detergents, such as anionic or cationic detergents, to yield microparticles with a surface having a net negative or a net positive charge. For example, microparticles manufactured with anionic detergents, such as hexadecyltrimethylammonium bromide (CTAB), i.e. CTAB-PLG microparticles, adsorb negatively charged macromolecules, such as DNA. (see, e.g., Int'l Application Number PCT/US99/17308).

Furthermore, other particulate systems and polymers can be used for the *in vivo* or *ex vivo* delivery of the gene of interest. For example, polymers such as polylysine, polyarginine, polyornithine, spermine, spermidine, as well as conjugates of these molecules, are useful for transferring a nucleic acid of interest. Similarly, DEAE dextran-mediated transfection, calcium phosphate precipitation or precipitation using other insoluble inorganic salts, such as strontium phosphate, aluminum silicates including bentonite and kaolin, chromic oxide, magnesium silicate, talc, and the like, will find use with the present methods. See, e.g., Felgner, P.L., *Advanced Drug Delivery Reviews* (1990) 5:163-187, for a review of delivery systems useful for gene transfer. Peptoids (Zuckerman, R.N., et al., U.S. Patent No. 5,831,005, issued November 3, 1998) may also be used for delivery of a construct of the present invention.

Additionally, biolistic delivery systems employing particulate carriers such as gold and tungsten, are especially useful for delivering synthetic expression cassettes

of the present invention. The particles are coated with the synthetic expression cassette(s) to be delivered and accelerated to high velocity, generally under a reduced atmosphere, using a gun powder discharge from a "gene gun." For a description of such techniques, and apparatuses useful therefore, see, e.g., U.S. Patent Nos. 4,945,050; 5,036,006; 5,100,792; 5,179,022; 5,371,015; and 5,478,744. Also, needleless injection systems can be used (Davis, H.L., et al, *Vaccine* 12:1503-1509, 1994; Bioject, Inc., Portland, OR).

Recombinant vectors carrying a synthetic expression cassette of the present invention are formulated into compositions for delivery to the vertebrate subject. These compositions may either be prophylactic (to prevent infection) or therapeutic (to treat disease after infection). The compositions will comprise a "therapeutically effective amount" of the gene of interest such that an amount of the antigen can be produced *in vivo* so that an immune response is generated in the individual to which it is administered. The exact amount necessary will vary depending on the subject being treated; the age and general condition of the subject to be treated; the capacity of the subject's immune system to synthesize antibodies; the degree of protection desired; the severity of the condition being treated; the particular antigen selected and its mode of administration, among other factors. An appropriate effective amount can be readily determined by one of skill in the art. Thus, a "therapeutically effective amount" will fall in a relatively broad range that can be determined through routine trials.

The compositions will generally include one or more "pharmaceutically acceptable excipients or vehicles" such as water, saline, glycerol, polyethyleneglycol, hyaluronic acid, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles. Certain facilitators of nucleic acid uptake and/or expression can also be included in the compositions or coadministered, such as, but not limited to, bupivacaine, cardiotoxin and sucrose.

Once formulated, the compositions of the invention can be administered directly to the subject (e.g., as described above) or, alternatively, delivered *ex vivo*, to cells derived from the subject, using methods such as those described above. For example, methods for the *ex vivo* delivery and reimplantation of transformed cells into

a subject are known in the art and can include, e.g., dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, lipofectamine and LT-1 mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) (with or without the corresponding antigen) in liposomes, and direct
5 microinjection of the DNA into nuclei.

Direct delivery of synthetic expression cassette compositions *in vivo* will generally be accomplished with or without viral vectors, as described above, by injection using either a conventional syringe or a gene gun, such as the Accell® gene delivery system (PowderJect Technologies, Inc., Oxford, England). The constructs
10 can be injected either subcutaneously, epidermally, intradermally, intramucosally such as nasally, rectally and vaginally, intraperitoneally, intravenously, orally or intramuscularly. Delivery of DNA into cells of the epidermis is particularly preferred as this mode of administration provides access to skin-associated lymphoid cells and provides for a transient presence of DNA in the recipient. Other modes of
15 administration include oral and pulmonary administration, suppositories, needle-less injection, transcutaneous and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. Administration of nucleic acids may also be combined with administration of peptides or other substances.

20 2.3.2 EX VIVO DELIVERY OF THE SYNTHETIC EXPRESSION CASSETTES OF THE PRESENT INVENTION

In one embodiment, T cells, and related cell types (including but not limited to antigen presenting cells, such as, macrophage, monocytes, lymphoid cells, dendritic cells, B-cells, T-cells, stem cells, and progenitor cells thereof), can be used for *ex vivo*
25 delivery of the synthetic expression cassettes of the present invention. T cells can be isolated from peripheral blood lymphocytes (PBLs) by a variety of procedures known to those skilled in the art. For example, T cell populations can be "enriched" from a population of PBLs through the removal of accessory and B cells. In particular, T cell enrichment can be accomplished by the elimination of non-T cells using anti-MHC
30 class II monoclonal antibodies. Similarly, other antibodies can be used to deplete specific populations of non-T cells. For example, anti-Ig antibody molecules can be

used to deplete B cells and anti-MacI antibody molecules can be used to deplete macrophages.

T cells can be further fractionated into a number of different subpopulations by techniques known to those skilled in the art. Two major subpopulations can be isolated based on their differential expression of the cell surface markers CD4 and CD8. For example, following the enrichment of T cells as described above, CD4⁺ cells can be enriched using antibodies specific for CD4 (see Coligan et al., *supra*). The antibodies may be coupled to a solid support such as magnetic beads. Conversely, CD8⁺ cells can be enriched through the use of antibodies specific for CD4 (to remove CD4⁺ cells), or can be isolated by the use of CD8 antibodies coupled to a solid support. CD4 lymphocytes from HIV-1 infected patients can be expanded *ex vivo*, before or after transduction as described by Wilson et. al. (1995) *J. Infect. Dis.* 172:88.

Following purification of T cells, a variety of methods of genetic modification known to those skilled in the art can be performed using non-viral or viral-based gene transfer vectors constructed as described herein. For example, one such approach involves transduction of the purified T cell population with vector-containing supernatant of cultures derived from vector producing cells. A second approach involves co-cultivation of an irradiated monolayer of vector-producing cells with the purified T cells. A third approach involves a similar co-cultivation approach; however, the purified T cells are pre-stimulated with various cytokines and cultured 48 hours prior to the co-cultivation with the irradiated vector producing cells. Pre-stimulation prior to such transduction increases effective gene transfer (Nolta et al. (1992) *Exp. Hematol.* 20:1065). Stimulation of these cultures to proliferate also provides increased cell populations for re-infusion into the patient. Subsequent to co-cultivation, T cells are collected from the vector producing cell monolayer, expanded, and frozen in liquid nitrogen.

Gene transfer vectors, containing one or more synthetic expression cassette of the present invention (associated with appropriate control elements for delivery to the isolated T cells) can be assembled using known methods.

Selectable markers can also be used in the construction of gene transfer vectors. For example, a marker can be used which imparts to a mammalian cell

transduced with the gene transfer vector resistance to a cytotoxic agent. The cytotoxic agent can be, but is not limited to, neomycin, aminoglycoside, tetracycline, chloramphenicol, sulfonamide, actinomycin, netropsin, distamycin A, anthracycline, or pyrazinamide. For example, neomycin phosphotransferase II imparts resistance to
5 the neomycin analogue geneticin (G418).

The T cells can also be maintained in a medium containing at least one type of growth factor prior to being selected. A variety of growth factors are known in the art which sustain the growth of a particular cell type. Examples of such growth factors are cytokine mitogens such as rIL-2, IL-10, IL-12, and IL-15, which promote growth
10 and activation of lymphocytes. Certain types of cells are stimulated by other growth factors such as hormones, including human chorionic gonadotropin (hCG) and human growth hormone. The selection of an appropriate growth factor for a particular cell population is readily accomplished by one of skill in the art.

For example, white blood cells such as differentiated progenitor and stem cells
15 are stimulated by a variety of growth factors. More particularly, IL-3, IL-4, IL-5, IL-6, IL-9, GM-CSF, M-CSF, and G-CSF, produced by activated T_H and activated macrophages, stimulate myeloid stem cells, which then differentiate into pluripotent stem cells, granulocyte-monocyte progenitors, eosinophil progenitors, basophil progenitors, megakaryocytes, and erythroid progenitors. Differentiation is modulated
20 by growth factors such as GM-CSF, IL-3, IL-6, IL-11, and EPO.

Pluripotent stem cells then differentiate into lymphoid stem cells, bone marrow stromal cells, T cell progenitors, B cell progenitors, thymocytes, T_H Cells, T_C cells, and B cells. This differentiation is modulated by growth factors such as IL-3, IL-4, IL-6, IL-7, GM-CSF, M-CSF, G-CSF, IL-2, and IL-5.

25 Granulocyte-monocyte progenitors differentiate to monocytes, macrophages, and neutrophils. Such differentiation is modulated by the growth factors GM-CSF, M-CSF, and IL-8. Eosinophil progenitors differentiate into eosinophils. This process is modulated by GM-CSF and IL-5.

The differentiation of basophil progenitors into mast cells and basophils is
30 modulated by GM-CSF, IL-4, and IL-9. Megakaryocytes produce platelets in response to GM-CSF, EPO, and IL-6. Erythroid progenitor cells differentiate into red blood cells in response to EPO.

Thus, during activation by the CD3-binding agent, T cells can also be contacted with a mitogen, for example a cytokine such as IL-2. In particularly preferred embodiments, the IL-2 is added to the population of T cells at a concentration of about 50 to 100 $\mu\text{g/ml}$. Activation with the CD3-binding agent can be carried out for 2 to 4 days.

Once suitably activated, the T cells are genetically modified by contacting the same with a suitable gene transfer vector under conditions that allow for transfection of the vectors into the T cells. Genetic modification is carried out when the cell density of the T cell population is between about 0.1×10^6 and 5×10^6 , preferably between about 0.5×10^6 and 2×10^6 . A number of suitable viral and nonviral-based gene transfer vectors have been described for use herein.

After transduction, transduced cells are selected away from non-transduced cells using known techniques. For example, if the gene transfer vector used in the transduction includes a selectable marker which confers resistance to a cytotoxic agent, the cells can be contacted with the appropriate cytotoxic agent, whereby non-transduced cells can be negatively selected away from the transduced cells. If the selectable marker is a cell surface marker, the cells can be contacted with a binding agent specific for the particular cell surface marker, whereby the transduced cells can be positively selected away from the population. The selection step can also entail fluorescence-activated cell sorting (FACS) techniques, such as where FACS is used to select cells from the population containing a particular surface marker, or the selection step can entail the use of magnetically responsive particles as retrievable supports for target cell capture and/or background removal.

More particularly, positive selection of the transduced cells can be performed using a FACS cell sorter (e.g. a FACSVantage™ Cell Sorter, Becton Dickinson Immunocytometry Systems, San Jose, CA) to sort and collect transduced cells expressing a selectable cell surface marker. Following transduction, the cells are stained with fluorescent-labeled antibody molecules directed against the particular cell surface marker. The amount of bound antibody on each cell can be measured by passing droplets containing the cells through the cell sorter. By imparting an electromagnetic charge to droplets containing the stained cells, the transduced cells can be separated from other cells. The positively selected cells are then harvested in

sterile collection vessels. These cell sorting procedures are described in detail, for example, in the FACSVantage™ Training Manual, with particular reference to sections 3-11 to 3-28 and 10-1 to 10-17.

Positive selection of the transduced cells can also be performed using
5 magnetic separation of cells based on expression of a particular cell surface marker. In such separation techniques, cells to be positively selected are first contacted with specific binding agent (e.g., an antibody or reagent that interacts specifically with the cell surface marker). The cells are then contacted with retrievable particles (e.g., magnetically responsive particles) which are coupled with a reagent that binds the
10 specific binding agent (that has bound to the positive cells). The cell-binding agent-particle complex can then be physically separated from non-labeled cells, for example using a magnetic field. When using magnetically responsive particles, the labeled cells can be retained in a container using a magnetic field while the negative cells are removed. These and similar separation procedures are known to those of ordinary
15 skill in the art.

Expression of the vector in the selected transduced cells can be assessed by a number of assays known to those skilled in the art. For example, Western blot or Northern analysis can be employed depending on the nature of the inserted nucleotide sequence of interest. Once expression has been established and the transformed T
20 cells have been tested for the presence of the selected synthetic expression cassette, they are ready for infusion into a patient via the peripheral blood stream.

The invention includes a kit for genetic modification of an *ex vivo* population of primary mammalian cells. The kit typically contains a gene transfer vector coding for at least one selectable marker and at least one synthetic expression cassette
25 contained in one or more containers, ancillary reagents or hardware, and instructions for use of the kit.

EXPERIMENTAL

Below are examples of specific embodiments for carrying out the present
30 invention. The examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way.

Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, temperatures, etc.), but some experimental error and deviation should, of course, be allowed for.

5

Example 1

Generation of Synthetic Expression Cassettes

A. Modification of HIV-1 *Env*, *Gag*, *Gag-protease* and *Gag-polymerase* Nucleic Acid Coding Sequences

10 The *Gag*, *Gag-protease*, and *Gag-polymerase* coding sequences were selected from the Type C strains AF110965 and AF110967. The *Env* coding sequences were selected from Type C strains AF110968 and AF110975. These sequences were manipulated to maximize expression of their gene products.

First, the HIV-1 codon usage pattern was modified so that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed
15 human genes. The HIV codon usage reflects a high content of the nucleotides A or T of the codon-triplet. The effect of the HIV-1 codon usage is a high AT content in the DNA sequence that results in a decreased translation ability and instability of the mRNA. In comparison, highly expressed human codons prefer the nucleotides G or C. The coding sequences were modified to be comparable to codon usage found in
20 highly expressed human genes.

Second, there are inhibitory (or instability) elements (INS) located within the coding sequences of the *Gag* and *Gag-protease* coding sequences (Schneider R, et al., *J Virol.* 71(7):4892-4903, 1997). RRE is a secondary RNA structure that interacts with the HIV encoded Rev-protein to overcome the expression down-regulating
25 effects of the INS. To overcome the post-transcriptional activating mechanisms of RRE and Rev, the instability elements are inactivated by introducing multiple point mutations that do not alter the reading frame of the encoded proteins. Figures 5 and 6 (SEQ ID Nos: 3, 4, 20 and 21) show the location of some remaining INS in synthetic sequences derived from strains AF110965 and AF110967. The changes made to these
30 sequences are boxed in the Figures. In Figures 5 and 6, the top line depicts a codon optimized sequence of *Gag* polypeptides from the indicated strains. The nucleotide(s) appearing below the line in the boxed region(s) depicts changes made to further

remove INS. Thus, when the changes indicated in the boxed regions are made, the resulting sequences correspond to the sequences depicted in Figures 1 and 2, respectively.

For the Gag-protease sequence, the changes in codon usage are restricted to the regions up to the -1 frameshift and starting again at the end of the Gag reading frame. Further, inhibitory (or instability) elements (INS) located within the coding sequences of the Gag-protease polypeptide coding sequence are altered as well. The synthetic coding sequences are assembled by methods known in the art, for example by companies such as the Midland Certified Reagent Company (Midland, Texas).

Modification of the Gag-polymerase sequences include similar modifications as described for Gag-protease in order to preserve the frameshift region.

In one embodiment of the invention, the full length polymerase coding region of the Gag-polymerase sequence is included with the synthetic Gag or Env sequences in order to increase the number of epitopes for virus-like particles expressed by the synthetic, optimized Gag/Env expression cassette. Because synthetic HIV-1 Gag-polymerase expresses the functional enzymes reverse transcriptase (RT) and integrase (INT) (in addition to the structural proteins and protease), it is important to inactivate RT and INT functions. Several deletions or mutations in the RT and INT coding regions can be made to achieve catalytic nonfunctional enzymes with respect to their RT and INT activity. {Jay. A. Levy (Editor) (1995) *The Retroviridae*, Plenum Press, New York. ISBN 0-306-45033X. Pages 215-20; Grimison, B. and Laurence, J. (1995), *Journal Of Acquired Immune Deficiency Syndromes and Human Retrovirology* 9(1):58-68; Wakefield, J. K., et al., (1992) *Journal Of Virology* 66(11):6806-6812; Esnouf, R., et al., (1995) *Nature Structural Biology* 2(4):303-308; Maignan, S., et al., (1998) *Journal Of Molecular Biology* 282(2):359-368; Katz, R. A. and Skalka, A. M. (1994) *Annual Review Of Biochemistry* 73 (1994); Jacobo-Molina, A., et al., (1993) *Proceedings Of the National Academy Of Sciences Of the United States Of America* 90(13):6320-6324; Hickman, A. B., et al., (1994) *Journal Of Biological Chemistry* 269(46):29279-29287; Goldgur, Y., et al., (1998) *Proceedings Of the National Academy Of Sciences Of the United States Of America* 95(16):9150-9154; Goette, M., et al., (1998) *Journal Of Biological Chemistry* 273(17):10139-10146; Gorton, J. L., et al., (1998) *Journal of Virology* 72(6):5046-5055; Engelman,

- A., et al., (1997) *Journal Of Virology* 71(5):3507-3514; Dyda, F., et al., *Science* 266(5193):1981-1986; Davies, J. F., et al., (1991) *Science* 252(5002):88-95; Bujacz, G., et al., (1996) *Febs Letters* 398(2-3):175-178; Beard, W. A., et al., (1996) *Journal Of Biological Chemistry* 271(21):12213-12220; Kohlstaedt, L. A., et al., (1992) *Science* 256(5065):1783-1790; Krug, M. S. and Berger, S. L. (1991) *Biochemistry* 30(44):10614-10623; Mazumder, A., et al., (1996) *Molecular Pharmacology* 49(4):621-628; Palaniappan, C., et al., (1997) *Journal Of Biological Chemistry* 272(17):11157-11164; Rodgers, D. W., et al., (1995) *Proceedings Of the National Academy Of Sciences Of the United States Of America* 92(4):1222-1226; Sheng, N. and Dennis, D. (1993) *Biochemistry* 32(18):4938-4942; Spence, R. A., et al., (1995) *Science* 267(5200):988-993.}

Furthermore selected B- and/or T-cell epitopes can be added to the Gag-polymerase constructs within the deletions of the RT- and INT-coding sequence to replace and augment any epitopes deleted by the functional modifications of RT and INT. Alternately, selected B- and T-cell epitopes (including CTL epitopes) from RT and INT can be included in a minimal VLP formed by expression of the synthetic Gag or synthetic GagProt cassette, described above. (For descriptions of known HIV B- and T-cell epitopes see, HIV Molecular Immunology Database CTL Search Interface; Los Alamos Sequence Compendia, 1987-1997; Internet address: <http://hiv-web.lanl.gov/immunology/index.html>.)

The resulting modified coding sequences are presented as a synthetic Env expression cassette; a synthetic Gag expression cassette; a synthetic Gag-protease expression cassette; and a synthetic Gag-polymerase expression cassette. A common Gag region (Gag-common) extends from nucleotide position 844 to position 903 (SEQ ID NO:1), relative to AF110965 (or from approximately amino acid residues 282 to 301 of SEQ ID NO:17) and from nucleotide position 841 to position 900 (SEQ ID NO:2), relative to AF110967 (or from approximately amino acid residues 281 to 300 of SEQ ID NO:22). A common Env region (Env-common) extends from nucleotide position 1213 to position 1353 (SEQ ID NO:5) and amino acid positions 405 to 451 of SEQ ID NO:23, relative to AF110968 and from nucleotide position 1210 to position 1353 (SEQ ID NO:11) and amino acid positions 404-451 (SEQ ID NO:24), relative to AF110975.

The synthetic DNA fragments for Gag and Env are cloned into the following eucaryotic expression vectors: pCMVKm2, for transient expression assays and DNA immunization studies, the pCMVKm2 vector is derived from pCMV6a (Chapman et al., *Nuc. Acids Res.* (1991) 19:3979-3986) and comprises a kanamycin selectable
5 marker, a ColE1 origin of replication, a CMV promoter enhancer and Intron A, followed by an insertion site for the synthetic sequences described below followed by a polyadenylation signal derived from bovine growth hormone -- the pCMVKm2 vector differs from the pCMV-link vector only in that a polylinker site is inserted into pCMVKm2 to generate pCMV-link; pESN2dhfr and pCMVPLEdhfr, for expression
10 in Chinese Hamster Ovary (CHO) cells; and, pAcC13, a shuttle vector for use in the Baculovirus expression system (pAcC13, is derived from pAcC12 which is described by Munemitsu S., et al., *Mol Cell Biol.* 10(11):5977-5982, 1990).

Briefly, construction of pCMVPLEdhfr was as follows.

To construct a DHFR cassette, the EMCV IRES (internal ribosome entry site)
15 leader was PCR-amplified from pCite-4a+ (Novagen, Inc., Milwaukee, WI) and inserted into pET-23d (Novagen, Inc., Milwaukee, WI) as an *Xba*-*Nco* fragment to give pET-EMCV. The *dhfr* gene was PCR-amplified from pESN2dhfr to give a product with a Gly-Gly-Gly-Ser spacer in place of the translation stop codon and inserted as an *Nco*-*Bam*H1 fragment to give pET-E-DHFR. Next, the attenuated *neo*
20 gene was PCR amplified from a pSV2Neo (Clontech, Palo Alto, CA) derivative and inserted into the unique *Bam*H1 site of pET-E-DHFR to give pET-E-DHFR/Neo_(m2). Finally the bovine growth hormone terminator from pCDNA3 (Invitrogen, Inc., Carlsbad, CA) was inserted downstream of the *neo* gene to give pET-E-DHFR/Neo_(m2)BGHt. The EMCV-*dhfr*/*neo* selectable marker cassette fragment was
25 prepared by cleavage of pET-E-DHFR/Neo_(m2)BGHt.

The CMV enhancer/promoter plus Intron A was transferred from pCMV6a (Chapman et al., *Nuc. Acids Res.* (1991) 19:3979-3986) as a *Hind*III-*Sal*I fragment into pUC19 (New England Biolabs, Inc., Beverly, MA). The vector backbone of pUC19 was deleted from the *Nde*I to the *Sap*I sites. The above described DHFR
30 cassette was added to the construct such that the EMCV IRES followed the CMV promoter. The vector also contained an amp^r gene and an SV40 origin of replication.

B. Defining of the Major Homology Region (MHR) of HIV-1 p55Gag

The Major Homology Region (MHR) of HIV-1 p55 (Gag) is located in the p24-CA sequence of Gag. It is a conserved stretch of approximately 20 amino acids. The position in the wild type AF110965 Gag protein is from 282-301 (SEQ ID NO:25) and spans a region from 844-903 (SEQ ID NO:26) for the Gag DNA-sequence. The position in the synthetic Gag protein is also from 282-301 (SEQ ID NO:25) and spans a region from 844-903 (SEQ ID NO:1) for the synthetic Gag DNA-sequence. The position in the wild type and synthetic AF110967 Gag protein is from 281-300 (SEQ ID NO:27) and spans a region from 841-900 (SEQ ID NO:2) for the modified Gag DNA-sequence. Mutations or deletions in the MHR can severely impair particle production (Borsetti, A., et al., *J. Virol.* 72(11):9313-9317, 1998; Mammano, F., et al., *J Virol* 68(8):4927-4936, 1994).

Percent identity to this sequence can be determined, for example, using the Smith-Waterman search algorithm (Time Logic, Incline Village, NV), with the following exemplary parameters: weight matrix = nuc4x4hb; gap opening penalty = 20, gap extension penalty = 5.

C. Defining of the Common Sequence Region of HIV-1 Env

The common sequence region (CSR) of HIV-1 Env is located in the C4 sequence of Env. It is a conserved stretch of approximately 47 amino acids. The position in the wild type and synthetic AF110968 Env protein is from approximately amino acid residue 405 to 451 (SEQ ID NO:28) and spans a region from 1213 to 1353 (SEQ ID NO:5) for the Env DNA-sequence. The position in the wild type and synthetic AF110975 Env protein is from approximately amino acid residue 404 to 451 (SEQ ID NO:29) and spans a region from 1210 to 1353 (SEQ ID NO:11) for the Env DNA-sequence.

Percent identity to this sequence can be determined, for example, using the Smith-Waterman search algorithm (Time Logic, Incline Village, NV), with the following exemplary parameters: weight matrix = nuc4x4hb; gap opening penalty = 20, gap extension penalty = 5.

Various forms of the different embodiments of the invention, described herein, may be combined.

Example 2

Expression Assays for the Synthetic Coding Sequences

A. Env, Gag and Gag-Protease Coding Sequences

5 The wild-type Env (from AF110968 or AF110975), Gag (from AF110965 and AF110967) and Gag-protease (from AF110965 and AF110967) sequences are cloned into expression vectors having the same features as the vectors into which the synthetic Env, Gag and Gag-protease sequences are cloned.

10 Expression efficiencies for various vectors carrying the wild-type and synthetic Env and Gag sequences are evaluated as follows. Cells from several mammalian cell lines (293, RD, COS-7, and CHO; all obtained from the American Type Culture Collection, 10801 University Boulevard, Manassas, VA 20110-2209) are transfected with 2 µg of DNA in transfection reagent LT1 (PanVera Corporation, 545 Science Dr., Madison, WI). The cells are incubated for 5 hours in reduced serum medium (Opti-MEM, Gibco-BRL, Gaithersburg, MD). The medium is then replaced with normal medium as follows: 293 cells, IMDM, 10% fetal calf serum, 2% glutamine (BioWhittaker, Walkersville, MD); RD and COS-7 cells, D-MEM, 10% fetal calf serum, 2% glutamine (Opti-MEM, Gibco-BRL, Gaithersburg, MD); and CHO cells, Ham's F-12, 10% fetal calf serum, 2% glutamine (Opti-MEM, Gibco-BRL, Gaithersburg, MD). The cells are incubated for either 48 or 60 hours. Cell lysates are collected as described below in Example 3. Supernatants are harvested and filtered through 0.45 µm syringe filters. Supernatants are evaluated using the Coulter p24-assay (Coulter Corporation, Hialeah, FL, US), using 96-well plates coated with a murine monoclonal antibody directed against HIV core antigen. The HIV-1 p24 antigen binds to the coated wells. Biotinylated antibodies against HIV recognize the bound p24 antigen. Conjugated streptavidin-horseradish peroxidase reacts with the biotin. Color develops from the reaction of peroxidase with TMB substrate. The reaction is terminated by addition of 4N H₂SO₄. The intensity of the color is directly proportional to the amount of HIV p24 antigen in a sample.

30 Synthetic Env, Gag and Gag-protease expression cassettes provides dramatic increases in production of their protein products, relative to the native (wild-type Type C) sequences, when expressed in a variety of cell lines.

Example 3

Western Blot Analysis of Expression

A. Env, Gag and Gag-Protease Coding Sequences

5 Human 293 cells are transfected as described in Example 2 with pCMV6a-based vectors containing native or synthetic Env or Gag expression cassettes. Cells are cultivated for 60 hours post-transfection. Supernatants are prepared as described. Cell lysates are prepared as follows. The cells are washed once with phosphate-buffered saline, lysed with detergent [1% NP40 (Sigma Chemical Co., St. Louis, MO)
10 in 0.1 M Tris-HCl, pH 7.5], and the lysate transferred into fresh tubes. SDS-polyacrylamide gels (pre-cast 8-16%; Novex, San Diego, CA) are loaded with 20 μ l of supernatant or 12.5 μ l of cell lysate. A protein standard is also loaded (5 μ l, broad size range standard; BioRad Laboratories, Hercules, CA). Electrophoresis is carried out and the proteins are transferred using a BioRad Transfer Chamber (BioRad
15 Laboratories, Hercules, CA) to Immobilon P membranes (Millipore Corp., Bedford, MA) using the transfer buffer recommended by the manufacturer (Millipore), where the transfer is performed at 100 volts for 90 minutes. The membranes are exposed to HIV-1-positive human patient serum and immunostained using o-phenylenediamine dihydrochloride (OPD; Sigma).
20 Immunoblotting analysis shows that cells containing the synthetic Env or Gag expression cassette produce the expected protein at higher per-cell concentrations than cells containing the native expression cassette. The proteins are seen in both cell lysates and supernatants. The levels of production are significantly higher in cell supernatants for cells transfected with the synthetic expression cassettes of the present
25 invention.

In addition, supernatants from the transfected 293 cells are fractionated on sucrose gradients. Aliquots of the supernatant are transferred to Polyclear™ ultracentrifuge tubes (Beckman Instruments, Columbia, MD), under-laid with a solution of 20% (wt/wt) sucrose, and subjected to 2 hours centrifugation at 28,000 rpm in a
30 Beckman SW28 rotor. The resulting pellet is suspended in PBS and layered onto a 20-60% (wt/wt) sucrose gradient and subjected to 2 hours centrifugation at 40,000 rpm in a Beckman SW41ti rotor.

The gradient is then fractionated into approximately 10 x 1 ml aliquots (starting at the top, 20%-end, of the gradient). Samples are taken from fractions 1-9 and are electrophoresed on 8-16% SDS polyacrylamide gels. The supernatants from 293/synthetic Env or Gag cells give much stronger bands than supernatants from 293/native Env or Gag cells.

Example 4

In Vivo Immunogenicity of Synthetic Gag and Env Expression Cassettes

A. Immunization

To evaluate the possibly improved immunogenicity of the synthetic Gag and Env expression cassettes, a mouse study is performed. The plasmid DNA, pCMVKM2 carrying the synthetic Gag expression cassette, is diluted to the following final concentrations in a total injection volume of 100 µl: 20 µg, 2 µg, 0.2 µg, 0.02 µg and 0.002 µg. To overcome possible negative dilution effects of the diluted DNA, the total DNA concentration in each sample is brought up to 20 µg using the vector (pCMVKM2) alone. As a control, plasmid DNA of the native Gag expression cassette is handled in the same manner. Twelve groups of four to ten Balb/c mice (Charles River, Boston, MA) are intramuscularly immunized (50 µl per leg, intramuscular injection into the *tibialis anterior*) according to the schedule in Table 1.

Table 1

Group	Gag or Env Expression Cassette	Concentration of Gag or Env plasmid DNA (μ g)	Immunized at time (weeks):
1	Synthetic	20	0 ¹ , 4
2	Synthetic	2	0, 4
3	Synthetic	0.2	0, 4
4	Synthetic	0.02	0, 4
5	Synthetic	0.002	0, 4
6	Synthetic	20	0
7	Synthetic	2	0
8	Synthetic	0.2	0
9	Synthetic	0.02	0
10	Synthetic	0.002	0
11	Native	20	0, 4
12	Native	2	0, 4
13	Native	0.2	0, 4
14	Native	0.02	0, 4
15	Native	0.002	0, 4
16	Native	20	0
17	Native	2	0
18	Native	0.2	0
19	Native	0.02	0
20	Native	0.002	0

¹ = initial immunization at "week 0"

Groups 1-5 and 11-15 are bled at week 0 (before immunization), week 4, week 6, week 8, and week 12. Groups 6-20 and 16-20 are bled at week 0 (before immunization) and at week 4.

B. Humoral Immune Response

The humoral immune response is checked with an anti-HIV Gag or Env antibody ELISAs (enzyme-linked immunosorbent assays) of the mice sera 0 and 4 weeks post immunization (groups 5-12) and, in addition, 6 and 8 weeks post immunization, respectively, 2 and 4 weeks post second immunization (groups 1-4).

The antibody titers of the sera are determined by anti-Gag or anti-Env antibody ELISA. Briefly, sera from immunized mice are screened for antibodies directed against the HIV p55 Gag protein or an Env protein, *e.g.* gp160 or gp120. ELISA microtiter plates are coated with 0.2 µg of Gag or Env protein per well overnight and washed four times; subsequently, blocking is done with PBS-0.2% Tween (Sigma) for 2 hours. After removal of the blocking solution, 100 µl of diluted mouse serum is added. Sera are tested at 1/25 dilutions and by serial 3-fold dilutions, thereafter. Microtiter plates are washed four times and incubated with a secondary, peroxidase-coupled anti-mouse IgG antibody (Pierce, Rockford, IL). ELISA plates are washed and 100 µl of 3, 3', 5, 5'-tetramethyl benzidine (TMB; Pierce) is added per well. The optical density of each well is measured after 15 minutes. The titers reported are the reciprocal of the dilution of serum that gave a half-maximum optical density (O.D.).

Synthetic expression cassettes will provide a clear improvement of immunogenicity relative to the native expression cassettes.

C. Cellular Immune Response

The frequency of specific cytotoxic T-lymphocytes (CTL) is evaluated by a standard chromium release assay of peptide pulsed Balb/c mouse CD4 cells. Gag or Env expressing vaccinia virus infected CD-8 cells are used as a positive control. Briefly, spleen cells (Effector cells, E) are obtained from the BALB/c mice immunized as described above are cultured, restimulated, and assayed for CTL activity against Gag peptide-pulsed target cells as described (Doe, B., and Walker, C.M., *AIDS* 10(7):793-794, 1996). Cytotoxic activity is measured in a standard ⁵¹Cr release assay. Target (T) cells are cultured with effector (E) cells at various E:T ratios for 4 hours and the average cpm from duplicate wells are used to calculate percent specific ⁵¹Cr release.

Cytotoxic T-cell (CTL) activity is measured in splenocytes recovered from the mice immunized with HIV Gag or Env DNA. Effector cells from the Gag or Env DNA-immunized animals exhibit specific lysis of Gag or Env peptide-pulsed SV-BALB (MHC matched) targets cells, indicative of a CTL response. Target cells that are peptide-pulsed and derived from an MHC-unmatched mouse strain (MC57) are not lysed.

Thus, synthetic Env and Gag expression cassettes exhibit increased potency for induction of cytotoxic T-lymphocyte (CTL) responses by DNA immunization.

10

Example 5

DNA-immunization of Non-Human Primates Using a Synthetic Env or Gag Expression Cassette

Non-human primates are immunized multiple times (*e.g.*, weeks 0, 4, 8 and 24) intradermally, mucosally or bilaterally, intramuscular, into the quadriceps using various doses (*e.g.*, 1-5 mg) synthetic Gag- and/or Env-containing plasmids. The animals are bled two weeks after each immunization and ELISA is performed with isolated plasma. The ELISA is performed essentially as described in Example 4 except the second antibody-conjugate is an anti-human IgG, g-chain specific, peroxidase conjugate (Sigma Chemical Co., St. Louis, MD 63178) used at a dilution of 1:500. Fifty µg/ml yeast extract is added to the dilutions of plasma samples and antibody conjugate to reduce non-specific background due to preexisting yeast antibodies in the non-human primates.

Further, lymphoproliferative responses to antigen can also be evaluated post-immunization, indicative of induction of T-helper cell functions.

Both synthetic Env and Gag plasmid DNA is expected to be immunogenic in non-human primates.

Example 6

In vitro expression of recombinant Sindbis RNA and DNA containing the synthetic Env and Gag expression cassette

30

To evaluate the expression efficiency of the synthetic Env and Gag expression cassette in Alphavirus vectors, the selected synthetic expression cassette is

subcloned into both plasmid DNA-based and recombinant vector particle-based Sindbis virus vectors. Specifically, a cDNA vector construct for *in vitro* transcription of Sindbis virus RNA vector replicons (pRSIN-luc; Dubensky, et al., *J Virol.* 70:508-519, 1996) is modified to contain a *PmeI* site for plasmid linearization and a polylinker for insertion of heterologous genes. A polylinker is generated using two oligonucleotides that contain the sites *XhoI*, *PmlI*, *ApaI*, *NarI*, *XbaI*, and *NotI* (XPANXNF, and XPANXNR).

The plasmid pRSIN-luc (Dubensky et al., *supra*) is digested with *XhoI* and *NotI* to remove the luciferase gene insert, blunt-ended using Klenow and dNTPs, and purified from an agarose gel using GeneCleanII (Bio101, Vista, CA). The oligonucleotides are annealed to each other and ligated into the plasmid. The resulting construct is digested with *NotI* and *SacI* to remove the minimal Sindbis 3'-end sequence and A₄₀ tract, and ligated with an approximately 0.4 kbp fragment from PKSSIN1-BV (WO 97/38087). This 0.4 kbp fragment is obtained by digestion of pKSSIN1-BV with *NotI* and *SacI*, and purification after size fractionation from an agarose gel. The fragment contains the complete Sindbis virus 3'-end, an A₄₀ tract and a *PmeI* site for linearization. This new vector construct is designated SINBVE.

The synthetic HIV Gag and Env coding sequences are obtained from the parental plasmid by digestion with *EcoRI*, blunt-ending with Klenow and dNTPs, purification with GeneCleanII, digestion with *SalI*, size fractionation on an agarose gel, and purification from the agarose gel using GeneCleanII. The synthetic Gag or Env coding fragment is ligated into the SINBVE vector that is digested with *XhoI* and *PmlI*. The resulting vector is purified using GeneCleanII and is designated SINBVGag. Vector RNA replicons may be transcribed *in vitro* (Dubensky et al., *supra*) from SINBVGag and used directly for transfection of cells. Alternatively, the replicons may be packaged into recombinant vector particles by co-transfection with defective helper RNAs or using an alphavirus packaging cell line.

The DNA-based Sindbis virus vector pDCMVSIN-beta-gal (Dubensky, et al., *J Virol.* 70:508-519, 1996) is digested with *SalI* and *XbaI*, to remove the beta-galactosidase gene insert, and purified using GeneCleanII after agarose gel size fractionation. The HIV Gag or Env gene is inserted into the the pDCMVSIN-beta-gal by digestion of SINBVGag with *SalI* and *XhoI*, purification using GeneCleanII of the

Gag-containing fragment after agarose gel size fractionation, and ligation. The resulting construct is designated pDSIN-Gag, and may be used directly for *in vivo* administration or formulated using any of the methods described herein.

5 BHK and 293 cells are transfected with recombinant Sindbis RNA and DNA, respectively. The supernatants and cell lysates are tested with the Coulter capture ELISA (Example 2).

BHK cells are transfected by electroporation with recombinant Sindbis RNA.

293 cells are transfected using LT-1 (Example 2) with recombinant Sindbis DNA. Synthetic Gag- and/or Env-containing plasmids are used as positive controls.
10 Supernatants and lysates are collected 48h post transfection.

Gag and Env proteins can be efficiently expressed from both DNA and RNA-based Sindbis vector systems using the synthetic expression cassettes.

Example 7

15 *In Vivo* Immunogenicity of recombinant Sindbis Replicon Vectors containing synthetic Gag and/or Env Expression Cassettes

A. Immunization

To evaluate the immunogenicity of recombinant synthetic Gag and Env expression cassettes in Sindbis replicons, a mouse study is performed. The Sindbis virus DNA vector carrying the synthetic Gag and/or Env expression cassette
20 (Example 6), is diluted to the following final concentrations in a total injection volume of 100 μ l: 20 μ g, 2 μ g, 0.2 μ g, 0.02 and 0.002 μ g. To overcome possible negative dilution effects of the diluted DNA, the total DNA concentration in each sample is brought up to 20 μ g using the Sindbis replicon vector DNA alone. Twelve
25 groups of four to ten Balb/c mice (Charles River, Boston, MA) are intramuscularly immunized (50 μ l per leg, intramuscular injection into the *tibialis anterior*) according to the schedule in Table 2. Alternatively, Sindbis viral particles are prepared at the following doses: 10^3 pfu, 10^5 pfu and 10^7 pfu in 100 μ l, as shown in Table 3. Sindbis Env or Gag particle preparations are administered to mice using intramuscular and
30 subcutaneous routes (50 μ l per site).

Table 2

Group	Gag or Env Expression Cassette	Concentration of Gag or Env DNA (μ g)	Immunized at time (weeks):
1	Synthetic	20	0 ¹ , 4
2	Synthetic	2	0, 4
3	Synthetic	0.2	0, 4
4	Synthetic	0.02	0, 4
5	Synthetic	0.002	0, 4
6	Synthetic	20	0
7	Synthetic	2	0
8	Synthetic	0.2	0
9	Synthetic	0.02	0
10	Synthetic	0.002	0

1 = initial immunization at "week 0"

Table 3

Group	Gag or Env sequence	Concentration of viral particle (pfu)	Immunized at time (weeks):
1	Synthetic	10 ³	0 ¹ , 4
2	Synthetic	10 ⁵	0, 4
3	Synthetic	10 ⁷	0, 4
8	Synthetic	10 ³	0
9	Synthetic	10 ⁵	0
10	Synthetic	10 ⁷	0

1 = initial immunization at "week 0"

Groups are bled and assessment of both humoral and cellular (e.g., frequency of specific CTLs) is performed, essentially as described in Example 4.

Although preferred embodiments of the subject invention have been described in some detail, it is understood that obvious variations can be made without departing from the spirit and the scope of the invention as defined by the appended claims.

CLAIMS

1. An expression cassette, comprising
a polynucleotide sequence encoding a polypeptide including an HIV *Gag*
5 polypeptide, wherein the polynucleotide sequence encoding said *Gag* polypeptide
comprises a sequence having at least 90% sequence identity to the sequence presented
as either nucleotides 844-903 of Figure 1 (SEQ ID NO:1) or nucleotides 841-900 of
Figure 2 (SEQ ID NO:2).
- 10 2. An expression cassette, comprising
a polynucleotide sequence encoding a polypeptide including an HIV *Gag*
polypeptide, wherein the polynucleotide sequence encoding said *Gag* polypeptide
comprises a sequence having at least 90% sequence identity to the sequence presented
as Figure 1 (SEQ ID NO:3) or Figure 2 (SEQ ID NO:4).
- 15 3. The expression cassette of claim 2, wherein said polynucleotide sequence
encoding a polypeptide including an HIV *Gag* polypeptide comprises a sequence
having at least 90% sequence identity to the sequence presented as Figure 1 (SEQ ID
NO:3).
- 20 4. The expression cassette of claim 2, wherein said polynucleotide sequence
encoding a polypeptide including an HIV *Gag* polypeptide comprises a sequence
having at least 90% sequence identity to the sequence presented as Figure 2 (SEQ ID
NO:4).
- 25 5. The expression cassette of claim 2, wherein the polynucleotide sequence
encoding said *Gag* polypeptide consists of a sequence having the sequence presented
as Figure 1 (SEQ ID NO:3).
- 30 6. The expression cassette of claim 2, wherein the polynucleotide sequence
encoding said *Gag* polypeptide consists of a sequence having the sequence presented
as Figure 2 (SEQ ID NO:4).

7. The expression cassette of any of claims 1 to 6, wherein said polynucleotide sequence further includes a polynucleotide sequence encoding an HIV *protease* polypeptide.

5

8. The expression cassette of any of claims 1 to 6, wherein said polynucleotide sequence further includes a polynucleotide sequence encoding an HIV *polymerase* polypeptide.

10

9. The expression cassette of any of claims 1 to 6, wherein said polynucleotide sequence further includes a polynucleotide sequence encoding an HIV *polymerase* polypeptide, wherein the sequence encoding the HIV *polymerase* polypeptide is modified by deletions of coding regions corresponding to reverse transcriptase and integrase.

15

10. The expression cassette of claim 9, wherein said polynucleotide sequence preserves T-helper cell and CTL epitopes.

20

11. An expression cassette, comprising a polynucleotide sequence encoding a polypeptide including an HIV *Env* polypeptide, wherein the polynucleotide sequence encoding said *Env* polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as nucleotides 1213-1353 of Figure 3 (SEQ ID NO:5).

25

12. The expression cassette of claim 11, wherein the polynucleotide sequence encoding said *Env* polypeptide further comprises a sequence having at least 90% sequence identity to the sequence presented as nucleotides 82-1512 of Figure 3 (SEQ ID NO:6).

30

13. The expression cassette of claim 11, wherein the polynucleotide sequence encoding said *Env* polypeptide further comprises a sequence having at least 90% sequence identity to the sequence presented as nucleotides 82-2025 of Figure 3

(SEQ ID NO:7).

14. The expression cassette of claim 11, wherein the polynucleotide sequence encoding said *Env* polypeptide further comprises a sequence having at least 90% sequence identity to the sequence presented as nucleotides 82-2547 of Figure 3 (SEQ ID NO:8).

15. The expression cassette of claim 11, wherein the polynucleotide sequence encoding said *Env* polypeptide further comprises a sequence having at least 90% sequence identity to the sequence presented as nucleotides 1-2547 of Figure 3 (SEQ ID NO:9).

16. An expression cassette, comprising a polynucleotide sequence encoding a polypeptide including an HIV *Env* polypeptide, wherein the polynucleotide sequence encoding said *Env* polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as nucleotides 1513-2547 of Figure 3 (SEQ ID NO:10).

17. An expression cassette, comprising a polynucleotide sequence encoding a polypeptide including an HIV *Env* polypeptide, wherein the polynucleotide sequence encoding said *Env* polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as nucleotides 1210-1353 of Figure 4 (SEQ ID NO:11).

18. The expression cassette of claim 17, wherein the polynucleotide sequence encoding said *Env* polypeptide further comprises a sequence having at least 90% sequence identity to the sequence presented as nucleotides 73-1509 of Figure 4 (SEQ ID NO:12).

19. The expression cassette of claim 17, wherein the polynucleotide sequence encoding said *Env* polypeptide further comprises a sequence having at least 90% sequence identity to the sequence presented as nucleotides 73-2022 of Figure 4

(SEQ ID NO:13).

20. The expression cassette of claim 17, wherein the polynucleotide sequence encoding said *Env* polypeptide further comprises a sequence having at least 90% sequence identity to the sequence presented as nucleotides 73-2565 of Figure 4 (SEQ ID NO:14).

21. The expression cassette of claim 17, wherein the polynucleotide sequence encoding said *Env* polypeptide further comprises a sequence having at least 90% sequence identity to the sequence presented as nucleotides 1-2565 of Figure 4 (SEQ ID NO:15).

22. An expression cassette, comprising a polynucleotide sequence encoding a polypeptide including an HIV *Env* polypeptide, wherein the polynucleotide sequence encoding said *Env* polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as nucleotides 1510-2565 of Figure 4 (SEQ ID NO:16).

23. An expression cassette, comprising a polynucleotide sequence encoding a polypeptide including an HIV *Env* polypeptide, wherein the polynucleotide sequence encoding said *Env* polypeptide consists of a sequence having the sequence presented as Figure 3 (SEQ ID NO:9) or Figure 4 (SEQ ID NO:15).

24. A recombinant expression system for use in a selected host cell, comprising, an expression cassette of any of claims 1-23, and wherein said polynucleotide sequence is operably linked to control elements compatible with expression in the selected host cell.

25. The recombinant expression system of claim 24, wherein said control elements are selected from the group consisting of a transcription promoter, a transcription enhancer element, a transcription termination signal, polyadenylation sequences, sequences for optimization of initiation of translation, and translation

termination sequences.

26. The recombinant expression system of claim 24, wherein said transcription promoter is selected from the group consisting of CMV, CMV+intron A, SV40, RSV, HIV-Ltr, MMLV-ltr, and metallothionein.

27. A cell comprising an expression cassette of any of claims 1-23, and wherein said polynucleotide sequence is operably linked to control elements compatible with expression in the selected cell.

28. The cell of claim 27, wherein the cell is a mammalian cell.

29. The cell of claim 28, wherein the cell is selected from the group consisting of BHK, VERO, HT1080, 293, RD, COS-7, and CHO cells.

30. The cell of claim 29, wherein said cell is a CHO cell.

31. The cell of claim 27, wherein the cell is an insect cell.

32. The cell of claim 31, wherein the cell is either *Trichoplusia ni* (Tn5) or Sf9 insect cells.

33. The cell of claim 27, wherein the cell is a bacterial cell.

34. The cell of claim 27, wherein the cell is a yeast cell.

35. The cell of claim 27, wherein the cell is a plant cell.

36. The cell of claim 27, wherein the cell is an antigen presenting cell.

37. The cell of claim 36, wherein the lymphoid cell is selected from the group consisting of macrophage, monocytes, dendritic cells, B-cells, T-cells, stem cells, and

progenitor cells thereof.

38. The cell of claim 27, wherein the cell is a primary cell.

5 39. The cell of claim 27, wherein the cell is an immortalized cell.

40. The cell of claim 27, wherein the cell is a tumor-derived cell.

10 41. A composition for generating an immunological response, comprising:
an expression cassette of any of claims 1-10.

42. The composition of claim 41, further comprising a *Gag* polypeptide
encoded by at least one of the expression cassettes of any of claims 1-10.

15 43. The composition of claim 41, further comprising an adjuvant.

44. A composition for generating an immunological response, comprising:
an expression cassette of any of claims 11-23.

20 45. The composition of claim 44, further comprising an *Env* polypeptide
encoded by at least one of the expression cassettes of any of claims 11-23.

46. The composition of either claim 44 or 45 further comprising:
an expression cassette of any of claims 1-10.

25 47. The composition of claim 46, further comprising the *Gag* polypeptide
encoded by at least one expression cassette of any of claims 1-10.

48. The composition of claim 44, further comprising an adjuvant.

30 49. A method of immunization of a subject, comprising,
introducing a composition of any of claims 41-48 into said subject under

conditions that are compatible with expression of said expression cassette in said subject.

5 50. The method of claim 49, wherein said expression cassette is introduced using a gene delivery vector.

 51. The method of claim 50, wherein the gene delivery vector is a non-viral vector.

10 52. The method of claim 50, wherein said gene delivery vector is a viral vector.

 53. The method of claim 52, wherein said gene delivery vector is a Sindbis-virus derived vector.

15 54. The method of claim 52, wherein said gene delivery vector is a retroviral vector.

 55. The method of claim 52, wherein said gene delivery vector is a lentiviral vector.

20 56. The method of claim 49, wherein said composition delivered using a particulate carrier.

25 57. The method of claim 49, wherein said composition is coated on a gold or tungsten particle and said coated particle is delivered to said subject using a gene gun.

 58. The method of claim 49, wherein said composition is encapsulated in a liposome preparation.

30 59. The method of any of claims 49-58, wherein said subject is a mammal.

60. The method of claim 59, wherein said mammal is a human.

61. A method of generating an immune response in a subject, comprising:
providing an expression cassette of any of claims 1-23,
5 expressing said polypeptide in a suitable host cell,
isolating said polypeptide, and
administering said polypeptide to the subject in an amount sufficient to elicit
an immune response.

10 62. A method of generating an immune response in a subject, comprising
introducing into cells of said subject an expression cassette of any of claims 1-
23, under conditions that permit the expression of said polynucleotide and production
of said polypeptide, thereby eliciting an immunological response to said polypeptide.

15 63. The method of claim 62, where the method further comprises
administration of a polypeptide produced by an expression cassette according to any
of claims 1-23 to the subject.

20 64. The method of claim 63, wherein administration of the polypeptide to the
subject is carried out before introducing said expression cassette.

65. The method of claim 63, wherein administration of the polypeptide to the
subject is carried out concurrently with introducing said expression cassette.

25 66. The method of claim 63, wherein administration of the polypeptide to the
subject is carried out after introducing said expression cassette.

1 / 6

Gag_AF110965_BW_mod

ATGGGCGCCCGCGCCAGCATCCTGCGCGGCGGCAAGCTGGACGCCTGGGAGCGCATCCGCC
TGCGCCCCGGCGGCAAGAAGTGCTACATGATGAAGCACCTGGTGTGGGCCAGCCGCGAGCT
GGAGAAGTTCGCCCTGAACCCCGGCTGCTGGAGACCAGCGAGGGCTGCAAGCAGATCATC
CGCCAGCTGCACCCCGCCCTGCAGACCGGCAGCGAGGAGCTGAAGAGCCTGTTCAACACCG
TGGCCACCCGTGTACTGCGTGACGAGAAGATCGAGGTCCGCGACACCAAGGAGGCCCTGGA
CAAGATCGAGGAGGAGCAGAACAAGTGCCAGCAGAAGATCCAGCAGGCCGAGGCCGCCGAC
AAGGGCAAGGTGAGCCAGAACTACCCCATCGTGCAGAACCTGCAGGGCCAGATGGTGCACC
AGGCCATCAGCCCCCGCACCCCTGAACGCCTGGGTGAAGGTGATCGAGGAGAAGGCCTTCAG
CCCCGAGGTGATCCCCATGTTACCGCCCTGAGCGAGGGCGCCACCCCCCAGGACCTGAAC
ACGATGTTGAACACCGTGGGCGGCCACCAGGCCGCCATGCAGATGCTGAAGGACACCATCA
ACGAGGAGGCCCGCCGAGTGGGACCGCGTGACCCCGTGACGCGCGGCCCATCGCCCCCGG
CCAGATGCGCGAGCCCCGCGGCAGCGACATCGCCGGCACCACCAGCACCCCTGCAGGAGCAG
ATCGCCTGGATGACCAGCAACCCCCCATCCCCGTGGGCGACATCTACAAGCGGTGGATCA
TCCTGGGCCTGAACAAGATCGTGCGGATGTACAGCCCCGTGAGCATCCTGGACATCAAGCA
GGGCCCCAAGGAGCCCTTCGCGACTACGTGGACCGCTTCTTCAAGACCCTGCGCGCCGAG
CAGAGCACCCAGGAGGTGAAGAACTGGATGACCGACACCCTGCTGGTGCAGAACGCCAACC
CCGACTGCAAGACCATCCTGCGCGCTCTCGGCCCCGGCGCCAGCCTGGAGGAGATGATGAC
CGCCTGCCAGGGCGTGGGCGGCCCCAGCCACAAGGCCCGCGTGCTGGCCGAGGCGATGAGC
CAGGCCAACACCAGCGTGATGATGCAGAAGAGCAACTTCAAGGGCCCCCGGCGCATCGTCA
AGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCGCCCCCGCAAGAA
GGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCC
AACTTCCTGGGCAAGATCTGGCCCAGCCACAAGGGCCGCCCGGCAACTTCCTGCAGAGCC
GCCCCGAGCCACCGCCCCCCCCCGCCGAGAGCTTCGCTTCGAGGAGACCACCCCGGCCA
GAAGCAGGAGAGCAAGGACCGCGAGACCCTGACCAGCCTGAAGAGCCTGTTTCGGCAACGAC
CCCCTGAGCCAGTAA

FIG. 1

2 / 6

Gag_AF110967_BW_mod

ATGGGCGCCCGCGCCAGCATCCTGCGCGGCGAGAAGCTGGACAAGTGGGAGAAGATCCGCC
TGCGCCCCGGCGGCAAGAAGCACTACATGCTGAAGCACCTGGTGTGGGCCAGCCGCGAGCT
GGAGGGCTTCGCCCTGAACCCCGGCCCTGCTGGAGACCGCCGAGGGCTGCAAGCAGATCATG
AAGCAGCTGCAGCCCGCCCTGCAGACCGGCACCGAGGAGCTGCGCAGCCTGTACAACACCG
TGGCCACCCCTGTACTGCGTGCACGCCGGCATCGAGGTCCGCGACACCAAGGAGGCCCTGGA
CAAGATCGAGGAGGAGCAGAACAAGTCCCAGCAGAAGACCCAGCAGGCCAAGGAGGCCGAC
GGCAAGGTGAGCCAGAACTACCCCATCGTGCAGAACCTGCAGGGCCAGATGGTGCACCAGG
CCATCAGCCCCCGCACCCCTGAACGCCTGGGTGAAGGTGATCGAGGAGAAGGCCTTCAGCCC
CGAGGTGATCCCCATGTTACCGCCCTGAGCGAGGGCGCCACCCCCAGGACCTGAACACG
ATGTTGAACACCGTGGGCGGCCACCAGGCCGCCATGCAGATGCTGAAGGACACCATCAACG
AGGAGGCCGCGGAGTGGGACCGCCTGCACCCCGTGCAGGCCGGCCCCGTTGGCCCCCGGCCA
GATGCGCGACCCCCGCGGCAGCGACATCGCCGGCGCCACCAGCACCTGCAGGAGCAGATC
GCCTGGATGACCAGCAACCCCCCGTGGCCGTGGGCGACATCTACAAGCGGTGGATCATCC
TGGGCCTGAACAAGATCGTGCGGATGTACAGCCCCGTGAGCATCCTGGACATCCGCCAGGG
CCCCAAGGAGCCCTTCGCGACTACGTGGACCGCTTCTTCAAGACCCTGCGCGCCGAGCAG
GCCACCCAGGACGTGAAGAACTGGATGACCGAGACCCTGCTGGTGCAGAACGCCAACCCCG
ACTGCAAGACCATCCTGCGCGCTCTCGGCCCCGGCGCCACCCTGGAGGAGATGATGACCGC
CTGCCAGGGCGTGGGCGGCCCCGGCCACAAGGCCCCGCGTGTGGCCGAGGCGATGAGCCAG
GCCAACAGCGTGAACATCATGATGCAGAAGAGCAACTTCAAGGGCCCCCGGCGCAACGTCA
AGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCAAGAACTGCCGCGCCCCCGCAAGAA
GGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCC
AACTTCCTGGGCAAGATCTGGCCCAGCCACAAGGGCCGCCCCGGCAACTTCCTGCAGAACC
GCAGCGAGCCCGCCGCCCCACCGTGCCCAACGCCCCCCCCCGCCGAGAGCTTCCGCTTCGA
GGAGACCACCCCGCCCCCAAGCAGGAGCCCAAGGACCGCGAGCCCTACCGCGAGCCCCTG
ACCGCCCTGCGCAGCCTGTTCCGCAGCGGCCCCCTGAGCCAGTAA

FIG. 2

3 / 6

Env_AF110968_C_BW_opt

--> signal peptide (1-81)
ATGCGCGTGATGGGCATCCTGAAGAACTACCAGCAGTGGTGGATGTGGGGCATCCTGGGCTTCTGGATGCTGATCA
TCAGCAGCGTGGTGGGCAACCTGTGGGTGACCGTGTACTACGGCGTGCCCGTGTGGAAGGAGGCCAAGACCACCT
GTTCGACCCAGCGACGCCAAGGCCTACGAGACCGAGGTGCACAACGTGTGGGCCACCCACGCTGCGTGCCCAACC
GACCCCAACCCCCAGGAGATCGTGCTGGAGAAGCTGACCGAGAACTTCAACATGTGGAAGAACGACATGGTGGACC
AGATGCACGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTGACCCCCCTGTGCGTGAC
CCTGAAGTGCCGCAACGTGAACGCCACCAACAACATCAACAGCATGATCGACAACAGCAACAAGGGCGAGATGAAG
AACTGCAGCTTCAACGTGACCACCGAGCTGCGCGACCGCAAGCAGGAGGTGCACGOCCTGTTCTACCGCCTGGACG
TGGTGCCCTGCGAGGGCAACAACAGCAACGAGTACCGCCTGATCAACTGCAACACCAGCGCCATCACCAGGCCTG
CCCCAAGGTGAGCTTCGACCCCATCCCCATCCACTACTGCACCCCGCCGGCTACGCCATCCTGAAGTGCAACAAC
CAGACCTTCAACGGCACCGGCCCCCTGCAACAACGTGAGCAGCGTGCAAGTGCGCCACGGCATCAAGCCCGTGGTGA
GCACCCAGTGCTGCTGAACGGCAGCCTGGCCAAGGGCGAGATCATCATCCGCGAGGAGAACCTGGCCAACAACGC
CAAGATCATCATCGTGACGTGAACAAGCCCGTGAAGATCGTGTGCGTGCGCCCAACAACAACACCCGCAAGAGC
GTGCGCATCGGCCCGGCCAGACCTTCTACGCCACCGCGAGATCATCGGCGACATCCGCCAGGCCTACTGCATCA
TCAACAAGACCGAGTGAACAGCACCTGCGAGGCGTGAGCAAGAAGCTGGAGGAGCACTTCAGCAAGAAGGCCAT
CAAGTTCGAGCCAGCAGCGGCGGACCTGGAGATCACCACCCACAGCTTCAACTGCCGCGCGAGTTCTTCTAC
TGCGACACCAGCCAGCTGTTCAACAGCACCTACAGCCCCAGCTTCAACGGCACCGAGAACAAGCTGAACGGCACCA
TCACCATCACTGCCGCATCAAGCAGATCATCAACATGTGGCAGAAGGTGGGCCGCGCCATGTACGCCCCCCCCAT
CGCCGGCAACCTGACCTGCGAGAGCAACATCACCAGCCTGCTGCTGACCCGCGACGGCGGCAAGACCGGCCCAAC
GACACCGAGATCTTCCGCCCCGGCGGCGGACATGCGCGACAACCTGGCGCAACGAGCTGTACAAGTACAAGGTGG
TGGAGATCAAGCCCTGGGCGTGGCCCCCACCAGGCGCAAGCGCCGCGTGGTGGAGCGCGAGAAGCGCGCCGTGGG
CATCGGCGCGTGTCTTCTGGGCTTCTGGGCGCGCGGCGAGCACCATGGGCGCGCCAGCATCACCCTGACCGTG
CAGGCCCCGCTGCTGCTGAGCGGCATCGTGACGAGCAGACAACCTGCTGCGCGCCATCGAGGCCAGCAGCACC
TGCTGCAGCTGACCGTGTGGGGCATCAAGCAGCTGCAGACCCGCATCCTGGCCGTGGAGCGCTACCTGAAGGACCA
GCAGCTGCTGGGCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCACCGCCGTGCCCTGGAACAGCAGCTGGAGC
AACCGCAGCCACGACGAGATCTGGGACAACATGACCTGGATGCAGTGGGACCGCGAGATCAACAACCTACACCGACA
CCATCTACCGCCTGCTGGAGGAGAGCCAGAACCAGCAGGAGAAGAACGAGAAGGACCTGCTGGCCCTGGACAGCTG
GCAGAACCTGTGGAACCTGTTTACGATCACCACCTGGCTGTGGTACATCAAGATCTTCATCATGATCGTGGGCGGC
CTGATCGGCCCTGCGCATCATCTTCGCCGTGCTGAGCATCGTGAACCGCGTGCGCCAGGGCTACAGCCCCCTGCCCT
TCCAGACCTGACCCCCAACCCCGGAGCCCGACCGCTGGGCCGATCGAGGAGGAGGGCGGCGAGCAGGACCG
CGGCCGAGCATCCGCTGCTGAGCGGCTTCTGGCCCTGGCCTGGGACGACCTGCGCAGCCTGTGCCTGTTTACG
TACCACCGCTGCGCGACTTCATCTGATCGCCGCCCGCGTGCTGGAGCTGCTGGGCCAGCGCGGCTGGGAGGCCC
TGAAGTACCTGGGCAGCCTGGTGCAGTACTGGGGCCTGGAGCTGAAGAAGAGCGCCATCAGCCTGCTGGACACCAT
CGCCATCGCCGTGGCCGAGGGCACCGACCGCATCATCGAGTTCATCCAGCGCATCTGCCGCGCCATCCGCAACATC
CCCCCGCATCCGCCAGGGCTTCGAGGCGCCCTGCAGTAA

FIG. 3

4 / 6

Env_AF110975_C_BW_opt

--> signal peptide (1-72) \ / -->
ATGCGCGTGGCGGCATCCTGCGCAGCTGGCAGCAGTGGTGGATCTGGGGCATCCTGGGCTTCTGGATCTGCAGCG
gp120/140/160 (72)
GCCTGGGCAACCTGTGGGTGACCGTGTACGACGGCGTGCCCGTGTGGCGCGAGGCCAGCACCACCCTGTTCTGCGC
CAGCGACGCCAAGGCCACGAGAAGGAGGTGCACAACGTGTGGGCCACCCACGCTGCGTGCCACCGACCCCAAC
CCCCAGGAGATCGAGCTGGACAACGTGACCGAGAATTCAACATGTGGAAGAAGCAGATGGTGGACCAGATGCACG
AGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCCGCGTGAAGCTGACCCCTGTGCGTGACCTGAAGTG
CACCAACTACAGCACCAACTACAGCAACACCATGAACGCCACCAGCTACAACAACAACACCACCGAGGAGATCAAG
AACTGCACCTTCAACATGACCACCGAGCTGCGCGACAAGAAGCAGCAGGTGTACGCCCTGTTCTACAAGCTGGACA
TCGTGCCCTTGAACAGCAACAGCAGCGAGTACCGCCTGATCAACTGCAACACCAGCGCCATCACCAGGCCCTGCC
CAAGGTGAGCTTCGACCCCATCCCCATCCACTACTGCGCCCCCGCGGCTACGCCATCCTGAAGTGCAAGAACAAC
ACCAGCAACGGCACCAGGCCCTGCCAGAACGTGAGCACCGTGCAGTGACCCACGGCATCAAGCCCGTGGTGAGCA
CCCCCTGCTGCTGAACGGCAGCCTGGCCGAGGGCGGCGAGATCATCATCCGACAGCAAGAACCTGAGCAACAACGC
CTACACCATCATCGTGACCTGAACGACAGCGTGGAGATCGTGTGACCCGCCCCAACACAACACCCGCAAGGGC
ATCCGCATCGGCCCCGGCCAGACCTTCTACGCCACCGAGAATCATTCGGCGACATCCGCCAGGCCCACTGCAACA
TCAGCGCCGGCGAGTGGAACAAGGCCGTGCAGCGCGTGAGCGCCAAGCTGCGCGAGCACTTCCCCAACAAGACCAT
CGAGTTCAGCCAGCAGCGGGCGGCGACCTGGAGATCACCACCCACAGCTTCAACTGCCGCGGCGAGTTCTTCTAC
TGCAACACCAGCAAGCTGTTCAACAGCAGCTACAACGGCACCAGCTACCGCGGCACCGAGAGCAACAGCAGCATCA
TCACCTGCCCCTGCCGCATCAAGCAGATCATCGACATGTGGCAGAAGGTGGGCGCGCCATCTACGCCCCCCCCAT
CGAGGGCAACATCACTGCAGCAGCAGCATCACCAGCCTGTGCTGGCCCGGACGGCGCCTGGACAACATCACC
ACCGAGATCTTCGCCCCCAGGGCGGCGACATGAAGGACAACCTGGCGCAACGAGCTGTACAAGTACAAGGTGGTGG
AGATCAAGCCCCCTGGGCGTGGCCCCCACCAGGCCAAGCGCGCGTGGTGGAGCGGAGAAGCGCGCGTGGGCAT
CGGCGCGGTGATCTTCGGCTTCTGGGCGCGCGGCGCAGCAACATGGGCGCGCCAGCATCACCCTGACCGCCAG
GCCCCCAGCTGCTGAGCGGCATCGTGCAGCAGCAGAGCAACCTGCTGCGGCCATCGAGGCCCAGCAGCACATGC
TGCAGCTGACCGTGTGGGGCATCAAGCAGCTGCAGGCCCGCGTGTGGCCATCGAGCGCTACCTGAAGGACCAGCA
GCTGTGGGCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCACCACCGTGCCTTGAACAGCAGCTGGAGCAAC
AAGACCCAGGGCGAGATCTGGGAGAACATGACCTGGATGCAGTGGGACAAGGAGATCAGCAACTACACCGGCATCA
TCTACCGCCTGCTGGAGGAGAGCCAGAACCAGCAGGAGCAGAACGAGAAGGACCTGCTGGCCCTGGACAGCCGCAA
CAACCTGTGGAGCTGGTTCAACATCAGCAACTGGCTGTGGTACATCAAGATCTTCATCATGATCGTGGGCGGCGCTG
ATCGGCCTGCGCATCATCTTCGCCGTGCTGAGCATCGTGAACCGCGTGCGCCAGGGCTACAGCCCCCTGAGCTTCC
AGACCCTGACCCCCAACCCCGCGGCCTGGACCGCTGGGCCGATCGAGGAGGAGGGCGGCGAGCAGGACCGCGA
CCGCAGCATCCGCCTGGTGCAGGGCTTCTGGCCCTGGCCTGGGACGACCTGCGCAGCCTGTGCCTGTTAGCTAC
CACCCTGCGCGACCTGATCCTGGTGACCGCCCGCGTGGTGGAGCTGCTGGGCGCAGCAGCCCCGCGGCCTGC
AGCGCGGCTGGGAGGCCCTGAAGTACCTGGGAGCCTGGTGCAGTACTGGGCGCTGGAGCTGAAGAAGAGCGCCAC
CAGCCTGCTGGACAGCATCGCCATCGCCGTGGCCGAGGGCAGCCGATCATCGAGGTGATCCAGCGCATCTAC
CGCGCCTTCTGCAACATCCCCCGCGCGTGGCCAGGGCTTCGAGGCCGCGCTGCAGTAA

FIG. 4

5 / 6

Gag_AF110965_BW_opt

ATGGGCGCCCGCGCCAGCATCCTGCGCGGCGGCAAGCTGGACGCCTGGGAGCGCATCCGCCTGCGCCCCGG
CGGCAAGAAGTGCTACATGATGAAGCACCTGGTGTGGGCCAGCCGCGAGCTGGAGAAGTTCGCCCTGAACC
CCGGCCTGCTGGAGACCAGCGAGGGCTGCAAGCAGATCATCCGCCAGCTGCACCCCGCCCTGCAGACCGGC
AGCGAGGAGCTGAAGAGCCTGTTCAACACCGTGGCCACCCTGTACTGCGTGCACGAGAAGATCGAGGTCCG
CGACACCAAGGAGGCCCTGGACAAGATCGAGGAGGAGCAGAACAAGTGCCAGCAGAAGATCCAGCAGGCCG
AGGCCGCCGACAAGGGCAAGGTGAGCCAGAACTACCCCATCGTGCAGAACCTGCAGGGCCAGATGGTGCAC
CAGGCCATCAGCCCCCGCACCTGAACGCCTGGGTGAAGGTGATCGAGGAGAAGGCCTTCAGCCCCGAGGT
GATCCCCATGTTACCGCCCTGAGCGAGGGCGCCACCCCCAGGACCTGAACAOCATGCTGAACACCGTGG
GCGGCCACCAGCGCCCATGCAGATGCTGAAGGACACCATCAACGAGGAGGCCCGCGAGTGGGACCGCGTG
CACCCCGTGCACGCCGGCCCCATCGCCCCGGCCAGATGCGCGAGCCCCGCGGCAGCGACATCGCCGGCAC
CACCAGCACCTGCAGGAGCAGATCGCCTGGATGACCAGCAACCCCCCATCCCCGTGGGCGACATCTACA
AGCGCTGGATCATCCTGGGCCTGAACAAGATCGTGCGCATGTACAGCCCCGTGAGCATCCTGGACATCAAG
CAGGGCCCCAAGGAGCCCTTCCGCGACTACGTGGACCGCTTCTTCAAGACCCTGCGCGCCGAGCAGAGCAC
CCAGGAGGTGAAGAACTGGATGACCGACACCCTGCTGGTGCAGAACGCCAACCCCGACTGCAAGACCATCC
TGCGCGCCTGGGCCCCGGCGCCAGCCTGGAGGAGATGATGACCGCCTGCCAGGGCGTGGGCGGCCCCAGC
CACAAGGCCCCGCGTGTGGCCGAGGCCATGAGCCAGGCCAACACCAGCGTGATGATGCAGAAGAGCAACTT
CAAGGGCCCCCGCGCATCGTGAAGTGCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCAACTGCCGCG
CCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAG
GCCAACTTCCTGGGCAAGATCTGGCCCAGCCACAAGGGCCGCCCGGCAACTTCCTGCAGAGCCGCCCCGA
GCCACCGCCCCCCCCCGCGAGAGCTTCCGCTTCGAGGAGACCACCCCGGCCAGAAGCAGGAGAGCAAGG
ACCGCGAGACCCTGACCAGCCTGAAGAGCCTGTTGCGCAACGACCCCTGAGCCAGTAA

FIG. 5

6 / 6

Gag_AF110967_BW_opt
ATGGGCGCCCGCCAGCATCCTGCGCGGCGAGAAGCTGGACAAGTGGGAGAAGATCCGCCCTGCGCCCCGG
CGGCAAGAAGCACTACATGCTGAAGCACCTGGTGTGGGCCAGCCGCGAGCTGGAGGGCTTCGCCCTGAACC
CCGGCCTGCTGGAGACCGCCGAGGGCTGCAAGCAGATCATGAAGCAGCTGCAGCCCGCCCTGCAGACCGGC
ACCGAGGAGCTGCGCAGCCTGTACAACACCGTGGCCACCCTGTACTGCGTGCACGCCGGCATCGAGGTGCG
CGACACCAAGGAGGGCCCTGGACAAGATCGAGGAGGAGCAGAACAAGAGCCAGCAGAAGACCCAGCAGGCCA
AGGAGGCCGACGGCAAGGTGAGCCAGAACTACCCCATCGTGCAGAACCTGCAGGGCCAGATGGTGCACCAG
GCCATCAGCCCCCGCACCCCTGAACGCCTGGGTGAAGGTGATCGAGGAGAAGGCCTTCAGCCCCGAGGTGAT
CCCCATGTTCAACGCCCTGAGCGAGGGCGCCACCCCCCAGGACCTGAACAACATGCTGAACACCGTGGGCG
GCCACCAGGCCGCCATGCAGATGCTGAAGGACACCATCAACGAGGAGGCCGCCGAGTGGGACCGCCTGCAC
CCCGTGCAGGCCCGCCCCGTGGCCCCCGGCCAGATGCGCGACCCCCGCGGCAGCGACATCGCCGGCGCCAC
CAGCACCCCTGCAGGAGCAGATCGCCTGGATGACCAGCAACCCCCCGTGCCCGTGGGCGACATCTACAAGC
GCTGGATCATCCTGGGCCTGAACAAGATCGTGCGCATGTACAGCCCCGTGAGCATCCTGGACATCCGCCAG
GGCCCCAAGGAGCCCTTCCGCGACTACGTGGACCGCTTCTTCAAGACCCTGCGCGCCGAGCAGGCCACCCA
GGACGTGAAGAACTGGATGACCGAGACCCTGCTGGTGCAGAACGCCAACCCGACTGCAAGACCATCCTGC
GCGCCCTGGGCCCCGGCGCCACCCTGGAGGAGATGATGACCGCCTGCCAGGGCGTGGGCGGCCCCGGCCAC
AAGGCCCCGCGTGTGGCCGAGGCCATGAGCCAGGCCAACAGCGTGAACATCATGATGCAGAAGAGCAACTT
CAAGGGCCCCCGCGCAACGTGAAGTGTCTCAACTGCGGCAAGGAGGGCCACATCGCCAAGAAGTGCCGCG
CCCCCGCAAGAAGGGCTGCTGGAAGTGGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAG
GCCAACTTCTGGGCAAGATCTGGCCAGCCACAAGGGCCGCCCCGGCAACTTCTGCAGAACCGCAGCGA
GCCCCGCGCCCCACCGTGCCACCGCCCCCCCCGCGAGAGCTTCCGCTTCGAGGAGACCACCCCCGCCC
CCAAGCAGGAGCCCAAGGACCGCGAGCCCTACCGCGAGCCCCTGACCGCCCTGCGCAGCCTGTTCCGCGAGC
GGCCCCCTGAGCCAGTAA

FIG. 6

SEQUENCE LISTING

<110> Chiron Corporation

<120> POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF

<130> 1631.100

<140>

<141>

<150> 60/152,195

<151> 1999-09-01

<160> 29

<170> PatentIn Ver. 2.0

<210> 1

<211> 60

<212> DNA

<213> Human immunodeficiency virus

<400> 1

gacatcaagc agggcccca ggagcccttc cgcgactacg tggaccgctt cttcaagacc 60

<210> 2

<211> 60

<212> DNA

<213> Human immunodeficiency virus

<400> 2

gacatccgcc agggcccca ggagcccttc cgcgactacg tggaccgctt cttcaagacc 60

<210> 3

<211> 1479

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic Gag
of HIV strain AF110965

<400> 3

atgggcgccc gcgccagcat cctgcgcggc ggcaagctgg acgcctggga gcgcatccgc 60
ctgcgccccg gcggcaagaa gtgctacatg atgaagcacc tgggtgtggc cagccgcgag 120
ctggagaagt tcgccctgaa ccccggcctg ctggagacca gcgagggctg caagcagatc 180
atccgccagc tgcacccgc cctgcagacc ggcagcgagg agctgaagag cctgttcaac 240
accgtggcca ccctgtactg cgtgcacgag aagatcgagg tccgcgacac caaggaggcc 300
ctggacaaga tcgaggagga gcagaacaag tgccagcaga agatccagca ggccgaggcc 360
gccgacaagg gcaagggtgag ccagaactac cccatcgtgc agaacctgca gggccagatg 420
gtgcaccagg ccatcagccc ccgcaccctg aacgcctggg tgaaggatgat cgaggagaag 480
gccttcagcc ccgaggtgat ccccatgttc accgcctga gcgagggcgc cccccccag 540
gacctgaaca cgatgttgaa caccgtgggc ggccaccagg ccgcatgca gatgctgaag 600
gacaccatca acgaggaggc cgccgagtgg gaccgcgtgc acccgtgca cgccggcccc 660
atcgcccccg gccagatgcg cgagcccgcc ggcagcgaca tcgcccggac caccagacc 720
ctgcaggagc agatcgccctg gatgaccagc aacccccca tccccgtggg cgacatctac 780

```

aagcgggtgga tcatcctggg cctgaacaag atcgtgcgga tgtacagccc cgtgagcatc 840
ctggacatca agcagggccc caaggagccc ttccgcgact acgtggaccg cttcttcaag 900
accctgcgcg ccgagcagag caccagggag gtgaagaact ggatgaccga caccctgctg 960
gtgcagaacg ccaaccccga ctgcaagacc atcctgcgcg ctctcggccc cggcgccagc 1020
ctggaggaga tgatgaccgc ctgccagggc gtgggcggcc ccagccacaa ggcccgcgtg 1080
ctggccgagg cgatgagcca ggccaacacc agcgtgatga tgcagaagag caacttcaag 1140
ggcccccgcc gcatcgtcaa gtgcttcaac tgcggcaagg agggccacat cgcccgaac 1200
tgccgcgccc cccgcaagaa gggctgctgg aagtgcggca aggagggcca ccagatgaag 1260
gactgcaccg agcgccaggc caacttctct ggcaagatct ggcccagcca caagggccgc 1320
cccggaact tcctgcagag ccgccccgag ccacccgccc ccccgccga gagcttccgc 1380
ttcgaggaga ccacccccgg ccagaagcag gagagcaagg accgcgagac cctgaccagc 1440
ctgaagagcc tggttcggca cgacccccctg agccagtaa 1479

```

<210> 4

<211> 1509

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic Gag
of HIV strain AF110967

<400> 4

```

atgggcgccc gcgccagcat cctgcgcggc gagaagctgg acaagtggga gaagatccgc 60
ctgcgccccg gcggcaagaa gcaactacatg ctgaagcacc tgggtgtgggc cagccgcgag 120
ctggagggct tcgccctgaa ccccgccctg ctggagaccg ccgagggctg caagcagatc 180
atgaagcagc tgcagcccgc cctgcagacc ggcaccgagg agctgcgcag cctgtacaac 240
accgtggcca cctgtactg cgtgcacgcc ggcatcgagg tccgcgacac caaggaggcc 300
ctggacaaga tcgaggagga gcagaacaag tcccagcaga agaccagca ggccaaggag 360
gccgacggca aggtgagcca gaactacccc atcgtgcaga acctgcaggg ccagatggtg 420
caccaggcca tcagcccccg caccctgaac gcctgggtga aggtgatcga ggagaaggcc 480
ttcagccccg aggtgatccc catgttcacc gccctgagcg agggcgccac ccccaggac 540
ctgaacacga tgttgaacac cgtgggcggc caccaggccg ccatgcagat gctgaaggac 600
accatcaacg aggaggccgc cgagtgggac cgcctgcacc ccgtgcaggc cggccccgtg 660
gcccccgccc agatgcgcga cccccgcggc agcgacatcg ccggcgccac cagcaccctg 720
caggagcaga tcgcctggat gaccagcaac cccccgtgc ccgtgggcga catctacaag 780
cggtggatca tcctgggcct gaacaagatc gtgcggatgt acagccccgt gagcatcctg 840
gacatccgcc agggccccaa ggagcccttc cgcgactacg tggaccgctt cttcaagacc 900
ctgcgcgcgc agcaggccac ccaggacgtg aagaactgga tgaccgagac cctgctggtg 960
cagaacgcca accccgactg caagaccatc ctgcgcgctc tcggccccgg cgccaccctg 1020
gaggagatga tgaccgcctg ccagggcgtg ggcggccccg gccacaaggc ccgcgtgctg 1080
gccgaggcga tgagccaggc caacagcgtg aacatcatga tgcagaagag caacttcaag 1140
ggcccccgcc gcaacgtcaa gtgcttcaac tgcggcaagg agggccacat cgccaagaac 1200
tgccgcgccc cccgcaagaa gggctgctgg aagtgcggca aggagggcca ccagatgaag 1260
gactgcaccg agcgccaggc caacttctct ggcaagatct ggcccagcca caagggccgc 1320
cccggaact tcctgcagaa ccgcagcag ccccgccccc ccaccgtgcc caccgcccc 1380
cccgccgaga gcttccgctt cgaggagacc acccccgc ccaagcagga gcccaaggac 1440
cgcgagccct accgcgagcc cctgaccgcc ctgcgcagcc tggttcggcag cggccccctg 1500
agccagtaa 1509

```

<210> 5

<211> 141

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Env common
region of HIV strain AF110968

<400> 5

accatcacca tcacctgccg catcaagcag atcatcaaca tgtggcagaa ggtggggccgc 60
 gccatgtacg ccccccccat cgccggcaac ctgacctgcg agagcaacat caccggcctg 120
 ctgctgacct gcgacggcgg c 141

<210> 6

<211> 1431

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic
 gp120 coding region of HIV strain AF110968

<400> 6

agcgtggtgg gcaacctgtg ggtgaccgtg tactacggcg tgcccgtgtg gaaggaggcc 60
 aagaccaccc tgttctgcac cagcgacgcc aaggcctacg agaccgaggt gcacaacctg 120
 tggggcaccac acgcctgcgt gccaccgcac cccaaccccc aggagatcgt gctggagaac 180
 gtgaccgaga acttcaacat gtggaagaac gacatggtgg accagatgca cgaggacatc 240
 atcagcctgt gggaccagag cctgaagccc tgcgtgaagc tgacccccct gtgctgacc 300
 ctgaagtgcc gcaacgtgaa cgccaccaac aacatcaaca gcatgatcga caacagcaac 360
 aagggcgaga tgaagaactg cagcttcaac gtgaccaccg agctgcgcga ccgcaagcag 420
 gaggtgcacg ccctgttcta cgcctggac gtggtgcccc tgcagggcaa caacagcaac 480
 gagtaccgcc tgatcaactg caacaccagc gccatcacc aggctgccc caaggtgagc 540
 ttcgacccca tccccatcca ctactgcacc cccgcggct acgccatcct gaagtgcac 600
 aaccagacct tcaacggcac cggccccctgc aacaacgtga gcagcgtgca gtgcgcccac 660
 ggcatacaagc ccgtggtgag caccagctg ctgctgaacg gcagcctggc caagggcgag 720
 atcatcatcc gcagcgagaa cctggccaac aacgccaaga tcatcatcgt gcagctgaac 780
 aagcccgtga agatcgtgtg cgtgcgcccc aacaacaaca cccgcaagag cgtgcgcac 840
 ggccccggcc agaccttcta cgccaccggc gagatcatcg gcgacatccg ccaggcctac 900
 tgcatacatca acaagaccga gtggaacagc accctgcagg gcgtgagcaa gaagctggag 960
 gagcacttca gcaagaaggc catcaagttc gagcccagca gcggcggcga cctggagatc 1020
 accaccacaca gcttcaactg ccgcggcgag ttcttctact gcgacaccag ccagctgttc 1080
 aacagcacct acagccccag cttcaacggc accgagaaca agctgaacgg caccatcacc 1140
 atcacctgcc gcatcaagca gatcatcaac atgtggcaga aggtgggccc cgccatgtac 1200
 gcccccccca tcgcccggcaa cctgacctgc gagagcaaca tcaccggcct gctgctgacc 1260
 cgcgacggcg gcaagaccgg ccccaacgac accgagatct tccgccccgg cggcgcgac 1320
 atgcgcgaca actggcgcaa cgagctgtac aagtacaagg tgggtggagat caagcccctg 1380
 ggcgtggccc ccaccgaggc caagcgccgc gtggtggagc gcgagaagcg c 1431

<210> 7

<211> 1944

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic
 gp140 coding region of HIV strain AF110968

<400> 7

agcgtggtgg gcaacctgtg ggtgaccgtg tactacggcg tgcccgtgtg gaaggaggcc 60
 aagaccaccc tgttctgcac cagcgacgcc aaggcctacg agaccgaggt gcacaacctg 120
 tggggcaccac acgcctgcgt gccaccgcac cccaaccccc aggagatcgt gctggagaac 180
 gtgaccgaga acttcaacat gtggaagaac gacatggtgg accagatgca cgaggacatc 240
 atcagcctgt gggaccagag cctgaagccc tgcgtgaagc tgacccccct gtgctgacc 300
 ctgaagtgcc gcaacgtgaa cgccaccaac aacatcaaca gcatgatcga caacagcaac 360
 aagggcgaga tgaagaactg cagcttcaac gtgaccaccg agctgcgcga ccgcaagcag 420
 gaggtgcacg ccctgttcta cgcctggac gtggtgcccc tgcagggcaa caacagcaac 480
 gagtaccgcc tgatcaactg caacaccagc gccatcacc aggctgccc caaggtgagc 540

```

ttcgacccca tccccatcca ctactgcacc cccgccggct acgccatcct gaagtgaac 600
aaccagacct tcaacggcac cggcccctgc aacaacgtga gcagcgtgca gtgcgcccac 660
ggcatcaagc ccgtggtgag caccagctg ctgctgaacg gcagcctggc caaggcgag 720
atcatcatcc gcagcgagaa cctggccaac aacgccaaga tcatcatcgt gcagctgaac 780
aagcccggtga agatcgtgtg cgtgcgcccc aacaacaaca cccgcaagag cgtgcgcatac 840
ggccccggcc agaccttcta cgccaccggc gagatcatcg gcgacatccg ccaggcctac 900
tgcacatcatc acaagaccga gtggaacagc accctgcagg gcgtgagcaa gaagctggag 960
gagcacttca gcaagaaggc catcaagttc gagcccagca gcggcgggcga cctggagatc 1020
accaccaca gcttcaactg ccgcggcgag ttcttttact gcgacaccag ccagctgttc 1080
aacagcacct acagccccag cttcaacggc accgagaaca agctgaacgg caccatcacc 1140
atcacctgcc gcatcaagca gatcatcaac atgtggcaga aggtgggccc cgccatgtac 1200
gcccccccca tcgccggcaa cctgacctgc gagagcaaca tcaccggcct gctgctgacc 1260
cgcgacggcg gcaagaccgg cccaacgac accgagatct tccgccccgg cggcgggcag 1320
atgcgcgaca actggcgcaa cgagctgtac aagtacaagg tgggtggagat caagcccctg 1380
ggcgtggccc ccaccgaggc caagcgccgc gtggtggagc gcgagaagcg cgcctggggc 1440
atcgggcgccg tggttcctggg cttcctgggc gccgcgggca gcacatggg cgccgccagc 1500
atcacctga ccgtgcaggc ccgcctgctg ctgagcggca tcgtgcagca gcagaacaac 1560
ctgctgcgcg ccacgaggc ccagcagcac ctgctgcagc tgaccgtgtg gggcatcaag 1620
cagctgcaga cccgcaccc cgcctggag cgctacctga aggaccagca gctgctgggc 1680
atctggggct gcagcgcaa gctgatctgc accaccgccc tgccctggaa cagcagctgg 1740
agcaaccgca gccacgacga gatctgggac aacatgacct ggatgcagtg ggaccggcag 1800
atcaacaact acaccgacac catctaccgc ctgctggagg agagccagaa ccagcaggag 1860
aagaacgaga aggacctgct ggccctggac agctggcaga acctgtggaa ctggttcagc 1920
atcaccaact ggctgtggta catc 1944

```

<210> 8

<211> 2466

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic
gp160 coding region of HIV strain AF110968

<400> 8

```

agcgtggtgg gcaacctgtg ggtgaccgtg tactacggcg tgcccgtgtg gaaggaggcc 60
aagaccaccc tgttctgcac cagcgacgcc aaggccctacg agaccgaggt gcacaacgtg 120
tgggcccacc acgcctgcgt gccaccgac cccaaccccc aggagatcgt gctggagaac 180
gtgaccgaga attcaacat gtggaagaac gacatggtgg accagatgca cgaggacatc 240
atcagcctgt gttgaccagag cctgaagccc tcgctgaagc tgacccccct gtgctgacc 300
ctgaagtgcc gcaacgtgaa cgccaccaac aacatcaaca gcatgatcga caacagcaac 360
aagggcgaga tgaagaactg cagcttcaac gtgaccaccg agctgcgcga ccgcaagcag 420
gaggtgcacg cctgtttcta ccgcctggac gtggtgcccc tgacgggcaa caacagcaac 480
gagtaccgcc tgatcaactg caacaccagc gccatcacc aggcctggcc caaggtgagc 540
ttcgacccca tccccatcca ctactgcacc cccgccggct acgccatcct gaagtgaac 600
aaccagacct tcaacggcac cggcccctgc aacaacgtga gcagcgtgca gtgcgcccac 660
ggcatcaagc ccgtggtgag caccagctg ctgctgaacg gcagcctggc caaggcgag 720
atcatcatcc gcagcgagaa cctggccaac aacgccaaga tcatcatcgt gcagctgaac 780
aagcccgtga agatcgtgtg cgtgcgcccc aacaacaaca cccgcaagag cgtgcgcatac 840
ggccccggcc agaccttcta cgccaccggc gagatcatcg gcgacatccg ccaggcctac 900
tgcacatcatc acaagaccga gtggaacagc accctgcagg gcgtgagcaa gaagctggag 960
gagcacttca gcaagaaggc catcaagttc gagcccagca gcggcgggcga cctggagatc 1020
accaccaca gcttcaactg ccgcggcgag ttcttttact gcgacaccag ccagctgttc 1080
aacagcacct acagccccag cttcaacggc accgagaaca agctgaacgg caccatcacc 1140
atcacctgcc gcatcaagca gatcatcaac atgtggcaga aggtgggccc cgccatgtac 1200
gcccccccca tcgccggcaa cctgacctgc gagagcaaca tcaccggcct gctgctgacc 1260
cgcgacggcg gcaagaccgg cccaacgac accgagatct tccgccccgg cggcgggcag 1320
atgcgcgaca actggcgcaa cgagctgtac aagtacaagg tgggtggagat caagcccctg 1380
ggcgtggccc ccaccgaggc caagcgccgc gtggtggagc gcgagaagcg cgcctggggc 1440

```

```

atcggcgcgcg tgttcctggg cttcctgggc gccgcgggca gcaccatggg cgccgccagc 1500
atcacctctga ccgtgcaggc ccgcctgctg ctgagcggca tcgtgcagca gcagaacaac 1560
ctgctgcgcg ccacgcaggc ccagcagcac ctgctgcagc tgaccgtgtg gggcatcaag 1620
cagctgcaga cccgcacccg ggccgtggag cgctacctga aggaccagca gctgctgggc 1680
atctggggct gcagcggcaa gctgatctgc accaccggcg tgccctggaa cagcagctgg 1740
agcaaccgca gccacgacga gatctgggac aacatgacct ggatgcagtg ggaccgcgag 1800
atcaacaact acaccgacac catctaccgc ctgctggagg agagccagaa ccagcaggag 1860
aagaacgaga aggacctgct ggccctggac agctggcaga acctgtggaa ctggttcagc 1920
atcaccaact ggctgtggta catcaagatc ttcatcatga tcgtgggcgg cctgatcggc 1980
ctgcgcatca tcttcgccgt gctgagcatc gtgaaccgcg tcgccagggg ctacagcccc 2040
ctgcccttcc agacctgac cccaacccc cgcgagccc accgcctggg ccgcatcgag 2100
gaggagggcg gcgagcagga ccgcggccgc agcatccgcc tgggtgagcg cttcctggcc 2160
ctggcctggg acgacctgag cagcctgtgc ctgttcagct accaccgcct gcgcgacttc 2220
atcctgatcg ccgcccgcgt gctggagctg ctgggcccagc gcggtggga ggccctgaag 2280
tacctgggca gcctgggtgca gtactggggc ctggagctga agaagagcgc catcagcctg 2340
ctggacacca tcgccatcgc cgtggccgag ggaccgacc gcacatcaga gttcatccag 2400
cgcatctgcc gcgccatccg caacatcccc cgccgcatcc gccagggctt cgaggccgcc 2460
ctgcag 2466

```

<210> 9

<211> 2547

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic
signal sequence and gp160 coding region of HIV
strain AF110968

<400> 9

```

atgcgcgtga tgggcatcct gaagaactac cagcagtggg ggatgtgggg catcctgggc 60
ttctggatgc tgatcatcag cagcgtgggt ggcaacctgt gggtagccgt gtactacggc 120
gtgcccgtgt ggaaggaggc caagaccacc ctgttctgca ccagcgacgc caaggccctac 180
gagaccgagg tgcacaacgt gtgggccacc cagcctgagc tgcccaccga cccaacccc 240
caggagatcg tgctggagaa cgtgaccgag aacttcaaca tgtggaagaa cgacatggtg 300
gaccagatgc acgaggacat catcagcctg tgggaccaga gcctgaagcc ctgctggaag 360
ctgaccccc tgtgcgtgac cctgaagtgc cgcaacgtga acgccaccaa caacatcaac 420
agcatgatcg acaacagcaa caagggcgag atgaagaact gcagcttcaa cgtgaccacc 480
gagctgcgcg accgcaagca ggaggtgcac gccctgttct accgcctgga cgtggtgccc 540
ctgcagggca acaacagcaa cgagtaccgc ctgatcaact gcaacaccag cgccatcacc 600
caggcctgcc ccaaggtgag cttcgacccc atccccatcc actactgcac ccccgccggc 660
tacgccatcc tgaagtgcaa caaccagacc ttcaacggca ccggcccctg caacaacgtg 720
agcagcgtgc agtgcccca cggcatcaag ccctggtgga gcaccagct gctgctgaac 780
ggcagcctgg ccaaggcgga gatcatcatc cgcagcgaga acctggccaa caacgccaag 840
atcatcatcg tgcagctgaa caagcccgtg aagatcgtgt gcgtgcgccc caacaacaac 900
acccgcaaga gcgtgcgcat cggccccggc cagaccttct acgccaccgg cgagatcatc 960
ggcgacatcc gccaggccta ctgcatcatc aacaagaccg agtggaacag caccctgcag 1020
ggcgtgagca agaagctgga ggagcacttc agcaagaagg ccatcaagt tgcagcccagc 1080
agcggcgggc acctggagat caccacccac agcttcaact gccgcggcga gttcttctac 1140
tgcgacacca gccagctgtt caacagcacc tacagcccca gcttcaacgg caccgagaac 1200
aagctgaacg gcacccatcac catcacctgc cgcacaaagc agatcatcaa catgtggcag 1260
aaggtggggc gcgccatgta cgcccccccc atcgccggca acctgacctg cgagagcaac 1320
atcaccggcc tgctgctgac ccgcgacggc ggcaagaccg gcccacacga caccgagatc 1380
ttccgcccgc gcggcgcgga catgcgcgac aactggcgca acgagctgta caagtacaag 1440
gtggtggaga tcaagccctt gggcgtggcc cccaccgagg ccaagcgccg cgtggtggag 1500
cgcgagaagc gcgcgtggg catcgccgac gtgttctctg gcttctctgg cgccgcccgc 1560
agcaccatgg gcgcgcccag catcacctg accgtgcagg cccgcctgct gctgagcggc 1620
atcgtgcagc agcagaacaa cctgctgcgc gccatcgagg cccagcagca cctgctgcag 1680
ctgaccgtgt ggggcatcaa gcagctgcag acccgcatcc tggcctgga gcgctacctg 1740

```

```

aaggaccagc agctgctggg catctggggc tgcagcggca agctgatctg caccaccgcc 1800
gtgccctgga acagcagctg gagcaaccgc agccacgacg agatctggga caacatgacc 1860
tggatgcagt gggaccgcga gatcaacaac tacaccgaca ccatctaccg cctgctggag 1920
gagagccaga accagcagga gaagaacgag aaggacctgc tggccctgga cagctggcag 1980
aacctgtgga actggttcag catcaccaac tggctgtggt acatcaagat cttcatcatg 2040
atcgtgggag gctgatcgg cctgcgcatc atcttcgccg tgctgagcat cgtgaaccgc 2100
gtgcgccagg gctacagccc cctgcccttc cagaccctga cccccaaccc ccgcgagccc 2160
gaccgcctgg gccgcacgga ggaggagggc ggcgagcagg accgcggccg cagcatccgc 2220
ctgggtgagc gcttcctggc cctggcctgg gacgacctgc gcagcctgtg cctgttcagc 2280
taccaccgcc tgcgcgactt catcctgacg gccgcccgcg tgctggagct gctgggcccag 2340
cgcggtctgg aggccctgaa gtacctgggc agcctggtgc agtactgggg cctggagctg 2400
aagaagagcg ccatcagcct gctggacacc atcgccatcg ccgtggccga gggcaccgac 2460
cgcacatcag agttcatcca gcgcacatgc cgcgccatcc gcaacatccc ccgccgcatc 2520
cgccagggct tcgaggccgc cctgcag 2547

```

<210> 10

<211> 1035

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic a
gp41 coding region of HIV strain AF110968

<400> 10

```

gccgtgggca tcggcgccgt gttcctgggc ttcctgggag ccgcggcag caccatgggc 60
gccgccagca tcaccctgac cgtgcaggcc cgcctgctgc tgagcggcat cgtgcagcag 120
cagaacaacc tgctgcgcgc catcgaggcc cagcagcacc tgctgcagct gaccgtgtgg 180
ggcatcaagc agctgcagac ccgcacctg gccgtggagc gctacctgaa ggaccagcag 240
ctgctgggca tctggggctg cagcggcaag ctgatctgca ccaccgccgt gccctggaac 300
agcagctgga gcaaccgcag ccacgacgag atctgggaca acatgacctg gatgcagtgg 360
gaccgcgaga tcaacaacta caccgacacc atctaccgcc tgctggagga gagccagaac 420
cagcaggaga agaacgagaa ggacctgctg gccctggaca gctggcagaa cctgtggaac 480
tgggttcagca tcaccaactg gctgtggtac atcaagatct tcatcatgat cgtgggcggc 540
ctgacgggcc tgcgcacatc cttcgccgtg ctgagcatcg tgaaccgcgt gcgccagggc 600
tacagccccc tgcccttcca gaccctgacc cccaaccccc gcgagcccga ccgcctgggc 660
cgcacgcagg aggaggcgcg cgagcaggac cgcgcccgca gcacccgcct ggtgagcggc 720
ttcctggccc tggcctggga cgacctgcgc agcctgtgcc tgttcagcta ccaccgcctg 780
cgcgacttca tctgatcgc cgcccgcgtg ctggagctgc tgggccagcg cggctgggag 840
gccctgaagt acctgggagc cctggtgcag tactggggcc tggagctgaa gaagagcgcc 900
atcagcctgc tggacaccat cgccatcgcc gtggccgagg gcaccgaccg catcatcgag 960
ttcatccagc gcacatgcgc cgccatccgc aacatcccc gccgcatccg ccagggcttc 1020
gaggccgccc tgcag 1035

```

<210> 11

<211> 144

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic Env
common region of HIV strain AF110975

<400> 11

```

agcatcatca ccctgccctg ccgcacaaag cagatcatcg acatgtggca gaaggtgggc 60
cgcgccatct acgccccccc catcgagggc aacatcacct gcagcagcag catcaccggc 120
ctgctgctgg ccgcgacgg cggc 144

```

<210> 12

<211> 1437

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic
gp120 coding region of HIV strain AF110975

<400> 12

```

agcggcctgg gcaacctgtg ggtgaccgtg tacgacggcg tgcccgtgtg gcgcgaggcc 60
agcaccaccc tgttctgcgc cagcgacgcc aaggcctacg agaaggaggt gcacaacgtg 120
tggggcacc cgccttgcgt gcccaccgac cccaaccccc aggagatcga gctggacaac 180
gtgaccgaga acttcaacat gtggaagaac gacatggtgg accagatgca cgaggacatc 240
atcagcctgt gggaccagag cctgaagccc cgcgtgaagc tgacccccct gtgcgtgacc 300
ctgaagtgc ccaactacag caccaactac agcaacacca tgaacgccac cagctacaac 360
aacaacacca ccgaggagat caagaactgc accttcaaca tgaccaccga gctgcgcgac 420
aagaagcagc aggtgtacgc cctgttctac aagctggaca tcgtgcccct gaacagcaac 480
agcagcgagt accgcctgat caactgcaac accagcgcca tcaccaggc ctgccccaa 540
gtgagcttcg accccatccc catccactac tgcgcccccg ccggctacgc catcctgaag 600
tgcaagaaca acaccagcaa cggcaccggc ccctgccaga acgtgagcac cgtgcagtgc 660
acccacggca tcaagcccgt ggtgagcacc cccctgctgc tgaacggcag cctggccgag 720
ggcggcgaga tcatcatccg cagcaagaac ctgagcaaca acgcctacac catcatcgtg 780
cacctgaacg acagcgtgga gatcgtgtgc acccgcccca acaacaacac ccgcaagggc 840
atccgcatcg gccccggcca gaccttctac gccaccgaga acatcatcgg cgacatccgc 900
caggcccact gcaacatcag cgccggcgag tggaacaagg ccgtgcagcg cgtgagcgcc 960
aagctgcgag agcaactccc caacaagacc atcgagttcc agcccagcag cggcggcgag 1020
ctggagatca ccaccacag cttcaactgc cgccggcgag tcttctactg caacaccagc 1080
aagctgttca acagcagcta caacggcacc agctaccgcg gcaccgagag caacagcagc 1140
atcatcaccc tgccctgccg catcaagcag atcatcgaca tgtggcagaa ggtgggcccgc 1200
gccatctacg ccccccccat cgaggggcaac atcacctgca gcagcagcat caccggcctg 1260
ctgctggccc gcgacggcgg cctggacaac atcaccaccg agatcttccg cccccagggc 1320
ggcgacatga aggacaactg gcgcaacgag ctgtacaagt acaagggtgt ggagatcaag 1380
cccctgggag tgccccccac cgaggccaag cgccgcgtgg tggagcgaga gaagcgc 1437

```

<210> 13

<211> 1950

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic
gp140 coding region of HIV strain AF110975

<400> 13

```

agcggcctgg gcaacctgtg ggtgaccgtg tacgacggcg tgcccgtgtg gcgcgaggcc 60
agcaccaccc tgttctgcgc cagcgacgcc aaggcctacg agaaggaggt gcacaacgtg 120
tggggcacc cgccttgcgt gcccaccgac cccaaccccc aggagatcga gctggacaac 180
gtgaccgaga acttcaacat gtggaagaac gacatggtgg accagatgca cgaggacatc 240
atcagcctgt gggaccagag cctgaagccc cgcgtgaagc tgacccccct gtgcgtgacc 300
ctgaagtgc ccaactacag caccaactac agcaacacca tgaacgccac cagctacaac 360
aacaacacca ccgaggagat caagaactgc accttcaaca tgaccaccga gctgcgcgac 420
aagaagcagc aggtgtacgc cctgttctac aagctggaca tcgtgcccct gaacagcaac 480
agcagcgagt accgcctgat caactgcaac accagcgcca tcaccaggc ctgccccaa 540
gtgagcttcg accccatccc catccactac tgcgcccccg ccggctacgc catcctgaag 600
tgcaagaaca acaccagcaa cggcaccggc ccctgccaga acgtgagcac cgtgcagtgc 660
acccacggca tcaagcccgt ggtgagcacc cccctgctgc tgaacggcag cctggccgag 720
ggcggcgaga tcatcatccg cagcaagaac ctgagcaaca acgcctacac catcatcgtg 780
cacctgaacg acagcgtgga gatcgtgtgc acccgcccca acaacaacac ccgcaagggc 840
atccgcatcg gccccggcca gaccttctac gccaccgaga acatcatcgg cgacatccgc 900

```

```

caggcccaact gcaacatcag cgccggcgag tggacaagcg cgtgcagcg cgtgagcgcc 960
aagctgcgcg agcacttccc caacaagacc atcgagttcc agcccagcag cggcggcgac 1020
ctggagatca ccaccacag cttcaactgc cgccggcgagt tcttctactg caacaccagc 1080
aagctgttca acagcagcta caacggcacc agctaccgcg gcaccgagag caacagcagc 1140
atcatcaccg tgccctgccg catcaagcag atcatcgaca tgtggcagaa ggtgggcccgc 1200
gccatctacg cccccccat cgaggggcaac atcacctgca gcagcagcat caccggcctg 1260
ctgctggccc gcgacggcgg cctggacaac atcaccaccg agatcttccg cccccagggc 1320
ggcgacatga aggacaactg gcgcaacgag ctgtacaagt acaagggtgt ggagatcaag 1380
cccctgggcg tggccccac cgaggccaag cgccgcgtgg tggagcgca gaagcgccg 1440
gtgggcatcg gcgcctgat cttcggttc ctggcgccc cggcagcaa catgggccc 1500
gccagcatca ccctgaccgc ccaggcccgc cagctgctga gcggcatcgt gcagcagcag 1560
agcaacctgc tgcgcgccat cgaggcccag cagcacatgc tgcagctgac cgtgtggggc 1620
atcaagcagc tgcaggcccg cgtgctggcc atcgagcgt acctgaagga ccagcagctg 1680
ctgggcatct ggggctgcag cggcaagctg atctgcacca ccaccgtgcc ctggaacagc 1740
agctggagca acaagaccca gggcgagatc tgggagaaca tgacctggat gcagtgggac 1800
aaggagatca gcaactacac cgcatcatc taccgctgc tggaggagag ccagaaccag 1860
caggagcaga acgagaagga cctgctggcc ctggacagcc gcaacaacct gtggagctgg 1920
ttcaacatca gcaactggct gtggtacatc 1950

```

<210> 14

<211> 2493

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic
gp160 coding region of HIV strain AF110975

<400> 14

```

agcggcctgg gcaacctgtg ggtgaccgtg tacgacggcg tgcccgtgtg gcgcgaggcc 60
agcaccaccc tgttctgcgc cagcgacgcc aaggcctacg agaaggaggt gcacaacgtg 120
tgggcccacc acgcctgcgt gccacccgac cccaaccccc aggagatcga gctggacaac 180
gtgaccgaga acttcaacat gtggaagaac gacatggtgg accagatgca cgaggacatc 240
atcagcctgt gggaccagag cctgaagccc cgctgaagc tgacccccct gtgctgacc 300
ctgaagtgca ccaactacag caccaactac agcaacacca tgaacgccac cagctacaac 360
aacaacacca ccgaggagat caagaactgc acctcaaca tgaccaccga gctgcgcgac 420
aagaagcagc aggtgtacgc cctgttctac aagctggaca tcgtgcccct gaacagcaac 480
agcagcgagt acccctgat caactgcaac accagcgcca tcaccaggc ctgccccaa 540
gtgagcttcg accccatccc catccactac tgcgcccccg ccggctacgc catcctgaag 600
tgcaagaaca acaccagcaa cggcaccggc ccctgccaga acgtgagcac cgtgcagtgc 660
acccacggca tcaagcccgt ggtgagcacc cccctgctgc tgaacggcag cctggccgag 720
ggcgggcaga tcatcatccg cagcaagaac ctgagcaaca acgcctacac catcatctg 780
cacctgaacg acagcgtgga gatcgtgtgc acccgcccca acaacaacac ccgcaagggc 840
atccgcatcg gcccgggcca gacctctac gccaccgaga acatcatcgg cgacatccg 900
caggcccaact gcaacatcag cgccggcgag tggacaagcg cgtgcagcg cgtgagcgcc 960
aagctgcgcg agcacttccc caacaagacc atcgagttcc agcccagcag cggcggcgac 1020
ctggagatca ccaccacag cttcaactgc cgccggcgagt tcttctactg caacaccagc 1080
aagctgttca acagcagcta caacggcacc agctaccgcg gcaccgagag caacagcagc 1140
atcatcaccg tgccctgccg catcaagcag atcatcgaca tgtggcagaa ggtgggcccgc 1200
gccatctacg cccccccat cgagggcaac atcacctgca gcagcagcat caccggcctg 1260
ctgctggccc gcgacggcgg cctggacaac atcaccaccg agatcttccg cccccagggc 1320
ggcgacatga aggacaactg gcgcaacgag ctgtacaagt acaagggtgt ggagatcaag 1380
cccctgggcg tggccccac cgaggccaag cgccgcgtgg tggagcgca gaagcgccg 1440
gtgggcatcg gcgcctgat cttcggttc ctggcgccc cggcagcaa catgggccc 1500
gccagcatca ccctgaccgc ccaggcccgc cagctgctga gcggcatcgt gcagcagcag 1560
agcaacctgc tgcgcgccat cgaggcccag cagcacatgc tgcagctgac cgtgtggggc 1620
atcaagcagc tgcaggcccg cgtgctggcc atcgagcgt acctgaagga ccagcagctg 1680
ctgggcatct ggggctgcag cggcaagctg atctgcacca ccaccgtgcc ctggaacagc 1740
agctggagca acaagaccca gggcgagatc tgggagaaca tgacctggat gcagtgggac 1800

```

```

aaggagatca gcaactacac cggcatcatc taccgcctgc tggaggagag ccagaaccag 1860
caggagcaga acgagaagga cctgctggcc ctggacagcc gcaacaacct gtggagctgg 1920
ttcaacatca gcaactggct gtggtacatc aagatcttca tcatgatcgt gggcggcctg 1980
atcgccctgc gcatcatctt cgccgtgctg agcatcgtga accgcgtgcg ccagggctac 2040
agccccctga gcttcagac cctgaccccc aacccccgcg gcctggaccg cctgggcccgc 2100
atcgaggagg agggcggcga gcaggaccgc gaccgcagca tccgcctggg gcagggcttc 2160
ctggccctgg cctgggacga cctgcgacgc ctgtgcctgt tcagctacca ccgcctgcgc 2220
gacctgatcc tggtgaccgc ccgcgtggtg gagctgctgg gccgcagcag cccccgcgcg 2280
ctgcagcgcg gctgggaggc cctgaagtac ctgggcagcc tggtgacgta ctggggcctg 2340
gagctgaaga agagcgccac cagcctgctg gacagcatcg ccatcgccgt ggccgagggc 2400
accgaccgca tcatcgaggt gatccagcgc atctaccgcg cttctgcaa catccccgcg 2460
cgcgctgcgc agggcttcga ggccgcctcg cag
2493

```

<210> 15

<211> 2565

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic
 signal sequence and gp160 coding region of HIV
 strain AF110975

<400> 15

```

atgcgcgtgc gcggcatcct gcgcagctgg cagcagtggg ggatctgggg catcctgggc 60
ttctggatct gcagcggcct gggcaacctg tgggtgaccg tgtacgacgg cgtgcccgtg 120
tggcgcgagg ccagcaccac cctgttctgc gccagcgacg ccaaggccta cgagaaggag 180
gtgcacaacg tgtggggcac ccacgcctgc gtgcccaccg accccaaccc ccaggagatc 240
gagctggaca acgtgaccga gaacttcaac atgtggaaga acgacatggg ggaccagatg 300
cacgaggaca tcatcagcct gtgggaccag agcctgaagc ccgcgtgaa gctgaccccc 360
ctgtgcgtga ccctgaagtg caccaactac agcaccaact acagcaaacac catgaacgcc 420
accagctaca acaacaacac caccgaggag atcaagaact gcaccttcaa catgaccacc 480
gagctgcgcg acaagaagca gcagggtgac gccctgttct acaagctgga catcgtgccc 540
ctgaacagca acagcagcga gtaccgcctg atcaactgca acaccagcgc catcacccag 600
gcctgcccc aaggtagctt cgacccccat cccatccact actgcgcccc cgccggctac 660
gccatcctga agtgcaagaa caacaccagc aacggcaccg gcccctgcca gaacgtgagc 720
accgtgcagt gcacccacgg catcaagccc gtggtgagca cccccctgct gctgaacggc 780
agcctggccg agggcggcga gatcatcatc cgacgcaaga acctgagcaa caacgcctac 840
accatcatcg tgcacctgaa cgacagcgtg gagatcgtgt gcacccgccc caacaacaac 900
acccgcaagg gcatccgcat cggccccggc cagaccttct acgccaccga gaacatcatc 960
ggcgacatcc gccaggccca ctgcaacatc agcgccggcg agtggaacaa ggccgtgcag 1020
cgcgtagcgc ccaagctgcg cgagcacttc ccaacaaga ccatcgagtt ccagcccagc 1080
agcgccggcg acctggagat caccaccac agcttcaact gccgcggcga gttcttctac 1140
tgcaacacca gcaagctgtt caacagcagc tacaacggca ccagctaccg cggcaccgag 1200
agcaacagca gcatcatcac cctgccttgc cgcataaagc agatcatcga catgtggcag 1260
aagggtgggc gcgccatcta cgccccccc atcgagggca acatcacctg cagcagcagc 1320
atcacccggc tgctgctggc ccgcgacggc ggctggaca acatcaccac cgagatcttc 1380
cgccccagg gcggcgacat gaaggacaac tggcgcaacg agctgtacaa gtacaagggtg 1440
gtggagatca agccccctgg cgtggccccc accgaggcca agcgccgcgt ggtggagcgc 1500
gagaagcgcg ccgtgggcat cggcgccgtg atcttcggct tcctgggcgc cgccggcagc 1560
aacatgggcg ccgccagcat caccctgacc gccaggccc gccagctgct gagcgcatc 1620
gtgcagcagc agagcaacct gctgcgcgcc atcgaggccc agcagcacat gctgcagctg 1680
accgtgtggg gcatcaagca gctgcaggcc cgctgctgg ccatcgagcg ctacctgaag 1740
gaccagcagc tgctgggcat ctggggctgc agcggaagc tgatctgcac caccaccgtg 1800
ccctggaaaca cgagctggag caacaagacc cagggcgaga tctgggagaa catgacctgg 1860
atgcagtggg acaaggagat cagcaactac accggcatca tctaccgctt gctggaggag 1920
agccagaacc agcaggagca gaacgagaag gacctgctgg ccctggacag ccgcaacaac 1980
ctgtggagct ggttcaacat cagcaactgg ctgtggtaca tcaagatctt catcatgatc 2040
gtgggcggcc tgatcggcct gcgcacatc ttcgcctgct tgagcatcgt gaaccgcgtg 2100

```

```

cgccagggct acagccccct gagcttccag accctgaccc ccaacccccg cggcctggac 2160
cgcttgggccc gcatcgagga ggagggcgcc gagcaggacc gcgaccgcag catccgcctg 2220
gtgcaggggt tcctggccct ggccctgggac gacctgcgca gcctgtgcct gttagctac 2280
caccgcctgc ggcacctgat cctggtgacc gcccgcgctg tggagctgct gggccgcagc 2340
agcccccgcg gcctgcagcg cggctgggag gccctgaagt acctgggcag cctggtgcag 2400
tactggggcc tggagctgaa gaagagcgcc accagcctgc tggacagcat cgccatcgcc 2460
gtggccgagg gcaccgaccg catcatcgag gtgatccagc gcatctaccg cgccttctgc 2520
aacatcccc gccgcgtgcg ccagggttc gagggcgccc tgcag 2565

```

<210> 16

<211> 1056

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic a
gp41 coding region of HIV strain AF110975

<400> 16

```

gccgtgggca tcggcgccgt gatcttcggc ttcttgggcg ccgcccggcag caacatgggc 60
gccgccagca tcaccctgac cgcccaggcc cgccagctgc tgagcggcat cgtgcagcag 120
cagagcaacc tgctgcgcgc catcgaggcc cagcagcaca tgctgcagct gaccgtgtgg 180
ggcatcaagc agctgcaggc ccgctgtgct gccatcgagc gctacctgaa ggaccagcag 240
ctgctgggca tctggggctg cagcggaag ctgatctgca ccaccaccgt gccctggaac 300
agcagctgga gcaacaagac ccagggcgag atctgggaga acatgacctg gatgcagtgg 360
gacaaggaga tcagcaacta caccggcatc atctaccgcc tgctggagga gagccagaac 420
cagcaggagc agaacagaaa ggacctgctg gccctggaca gccgcaacaa cctgtggagc 480
tggttcaaca tcagcaactg gctgtggtac atcaagatct tcatcatgat cgtgggcggc 540
ctgatcgccc tgccgcatcat cttcgccgtg ctgagcatcg tgaaccgcgt gcgccagggc 600
tacagcccc tgagcttcca gacctgacc cccaaccccc gcggcctgga ccgcctgggc 660
cgcatcgagg aggaggcgcg cgagcaggac cgcgaccgca gcatccgcct ggtgcagggc 720
ttcttggccc tggcctggga cgacctgcgc agcctgtgcc tggtagcta ccaccgcctg 780
cgcgacctga tcctggtgac cgcccgctg gtggagctgc tgggcccag cagccccgc 840
ggcctgcagc gcggctggga ggccctgaag tacctgggca gcctggtgca gtactggggc 900
ctggagctga agaagagcgc caccagcctg ctggacagca tcgccatcg cgtggccgag 960
ggcaccgacc gcatcatcga ggtgatccag cgcatctacc gcgccttctg caacatcccc 1020
cgccgcgtgc gccagggttc cgaggccgccc ctgcag 1056

```

<210> 17

<211> 492

<212> PRT

<213> Human immunodeficiency virus

<400> 17

```

Met Gly Ala Arg Ala Ser Ile Leu Arg Gly Gly Lys Leu Asp Ala Trp
  1              5              10              15

Glu Arg Ile Arg Leu Arg Pro Gly Gly Lys Lys Cys Tyr Met Met Lys
      20              25              30

His Leu Val Trp Ala Ser Arg Glu Leu Glu Lys Phe Ala Leu Asn Pro
      35              40              45

Gly Leu Leu Glu Thr Ser Glu Gly Cys Lys Gln Ile Ile Arg Gln Leu
      50              55              60

His Pro Ala Leu Gln Thr Gly Ser Glu Glu Leu Lys Ser Leu Phe Asn
      65              70              75              80

```


Thr Val Ala Thr Leu Tyr Cys Val His Glu Lys Ile Glu Val Arg Asp
 85 90 95
 Thr Lys Glu Ala Leu Asp Lys Ile Glu Glu Glu Gln Asn Lys Cys Gln
 100 105 110
 Gln Lys Ile Gln Gln Ala Glu Ala Ala Asp Lys Gly Lys Val Ser Gln
 115 120 125
 Asn Tyr Pro Ile Val Gln Asn Leu Gln Gly Gln Met Val His Gln Ala
 130 135 140
 Ile Ser Pro Arg Thr Leu Asn Ala Trp Val Lys Val Ile Glu Glu Lys
 145 150 155 160
 Ala Phe Ser Pro Glu Val Ile Pro Met Phe Thr Ala Leu Ser Glu Gly
 165 170 175
 Ala Thr Pro Gln Asp Leu Asn Thr Met Leu Asn Thr Val Gly Gly His
 180 185 190
 Gln Ala Ala Met Gln Met Leu Lys Asp Thr Ile Asn Glu Glu Ala Ala
 195 200 205
 Glu Trp Asp Arg Val His Pro Val His Ala Gly Pro Ile Ala Pro Gly
 210 215 220
 Gln Met Arg Glu Pro Arg Gly Ser Asp Ile Ala Gly Thr Thr Ser Thr
 225 230 235 240
 Leu Gln Glu Gln Ile Ala Trp Met Thr Ser Asn Pro Pro Ile Pro Val
 245 250 255
 Gly Asp Ile Tyr Lys Arg Trp Ile Ile Leu Gly Leu Asn Lys Ile Val
 260 265 270
 Arg Met Tyr Ser Pro Val Ser Ile Leu Asp Ile Lys Gln Gly Pro Lys
 275 280 285
 Glu Pro Phe Arg Asp Tyr Val Asp Arg Phe Phe Lys Thr Leu Arg Ala
 290 295 300
 Glu Gln Ser Thr Gln Glu Val Lys Asn Trp Met Thr Asp Thr Leu Leu
 305 310 315 320
 Val Gln Asn Ala Asn Pro Asp Cys Lys Thr Ile Leu Arg Ala Leu Gly
 325 330 335
 Pro Gly Ala Ser Leu Glu Glu Met Met Thr Ala Cys Gln Gly Val Gly
 340 345 350
 Gly Pro Ser His Lys Ala Arg Val Leu Ala Glu Ala Met Ser Gln Ala
 355 360 365
 Asn Thr Ser Val Met Met Gln Lys Ser Asn Phe Lys Gly Pro Arg Arg
 370 375 380
 Ile Val Lys Cys Phe Asn Cys Gly Lys Glu Gly His Ile Ala Arg Asn
 385 390 395 400

```
<400> 20
atggggcgccc  gcgccagcat  cctgcgcggc  ggcaagctgg  acgcctggga  ggcgatccgc  60
ctggcgcccc  cgggcaagaa  gtgctacatg  atgaagcacc  tgggtgtggc  cagccgcgag  120
ctggagaagt  tgcgcttgaa  ccccgcgctg  ctggagacca  gcgagggctg  caagcatgat  180
atccgccagc  tgcaccccg  cctgcagacc  ggcagcgagg  agctgaagag  cctgttcaac  240
```

```

accgtggcca ccctgtactg cgtgcacgag aagatcgagg tgcgcgacac caaggaggcc 300
ctggacaaga tcgaggagga gcagaacaag tgccagcaga agatccagca ggccgaggcc 360
gccgacaagg gcaagggtgag ccagaactac cccatcgtgc agaacctgca gggccagatg 420
gtgcaccagg ccatcagccc ccgcacctg aacgcctggg tgaagggtgat cgaggagaag 480
gccttcagcc ccgagggtgat ccccatgttc accgcctga gcgagggcgc cccccccag 540
gacctgaaca ccatgtgtaa caccgtgggc ggccaccagg ccgccatgca gatgctgaag 600
gacaccatca acgaggaggc cgccgagtgg gaccgcgtgc acccctgca cgccggcccc 660
atcgcccccg gccagatgcg cgagccccgc ggcagcgaca tcgccggcac caccagcacc 720
ctgcaggagc agatcgcttg gatgaccagc aacccccca tccccgtggg cgacatctac 780
aagcgctgga tcctcctggg cctgaacaag atcgtgcgca tgtacagccc cgtgagcatc 840
ctggacatca agcaggggccc caaggaggccc ttccgcgact acgtggaccg cttcttcaag 900
accctgcgcg ccgagcagag caccaggagg gtgaagaact ggatgaccga caccctgctg 960
gtgcagaacg ccaacccccg ctgcaagacc atcctgcgcg ccctgggccc cgccgccagc 1020
ctggaggaga tgatgaccgc ctgccagggc gtgggcggcc ccagccacaa ggcccgcgtg 1080
ctggccgagg ccatgagcca ggccaacacc agcgtgatga tgcagaagag caacttcaag 1140
ggcccccgcc gcatcgtgaa gtgcttcaac tgcggcaagg agggccacat cgcccgaac 1200
tgccgcgccc cccgcaagaa gggctgctgg aagtgcggca aggagggcca ccagatgaag 1260
gactgcaccg agcgccaggc caacttcctg ggcaagatct ggcccagcca caaggggccg 1320
cccggcaact tcctgcagag ccgccccgag cccaccgccc cccccccga gagcttccg 1380
ttcgaggaga ccacccccg ccagaagcag gagagcaagg accgcgagac cctgaccagc 1440
ctgaagagcc tgttcggcaa cgacccccctg agccagtaa 1479

```

<210> 21

<211> 1509

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic Gag
coding sequence of HIV strain AF110967

<400> 21

```

atgggcgccc gcgccagcat cctgcgcggc gagaagctgg acaagtggga gaagatccgc 60
ctgcgccccg gcggcaagaa gcactacatg ctgaagcacc tgggtgtggc cagccgcgag 120
ctggaggggt tcgcctgaa ccccgccctg ctggagaccg ccgagggctg caagcagatc 180
atgaagcagc tgcagcccg cctgcagacc ggcaccgagg agctgcgcag cctgtacaac 240
accgtggcca ccctgtactg cgtgcacgcc ggcacgagg tgcgcgacac caaggaggcc 300
ctggacaaga tcgaggagga gcagaacaag agccagcaga agaccagca ggccaaggag 360
gccgacggca aggtgagcca gaactacccc atcgtgcaga acctgcagg ccagatggtg 420
caccaggcca tcagccccc caccctgaac gcctgggtga aggtgatcga ggagaaggcc 480
ttcagccccg aggtgatccc catgttcacc gccctgagcg agggcgccac ccccaggac 540
ctgaacacca tgctgaacac cgtgggcggc caccaggccg ccatgcagat gctgaaggac 600
accatcaacg aggaggccgc cgagtgggac cgcctgcacc ccgtgcagge cgccccctg 660
gcccccgccc agatgcgcga cccccgcggc agcgacatcg ccggcgccac cagcaccctg 720
caggagcaga tcgcctggat gaccagcaac cccccctg cctggggcga catctacaag 780
cgctggatca tcctgggcct gaacaagatc gtgcgcgatg acagccccgt gagcatcctg 840
gacatccgcc agggccccc ggagcccttc cgcgactacg tggaccgctt cttcaagacc 900
ctgcgcgcg agcaggccac ccaggacgtg aagaactgga tgaccgagac cctgctggtg 960
cagaacgcca accccgactg caagaccatc ctgcgcgccc tgggcccccg cgccaccctg 1020
gaggagatga tgaccgcctg ccagggcctg ggcggccccc gccacaagg ccgcgtgctg 1080
gccgaggcca tgagccaggc caacagcgtg aacatcatga tgcagaagag caacttcaag 1140
ggcccccgcc gcaacgtgaa gtgcttcaac tgcggcaagg agggccacat cgccaagaac 1200
tgccgcgccc cccgcaagaa gggctgctgg aagtgcggca aggagggcca ccagatgaag 1260
gactgcaccg agcgccaggc caacttcctg ggcaagatct ggcccagcca caaggggccg 1320
cccggcaact tcctgcagaa ccgcagcag cccgcgcgcc ccaccgtgcc caccgcccc 1380
cccggcgaga gcttcgcctt cgaggagacc acccccgc ccaagcagga gccaaggac 1440
cgcgagccct accgcgagcc cctgaccgcc ctgcgcagcc tgttcggcag cgccccctg 1500
agccagtaa 1509

```

<210> 22

<211> 502

<212> PRT

<213> Human immunodeficiency virus

<400> 22

```

Met Gly Ala Arg Ala Ser Ile Leu Arg Gly Glu Lys Leu Asp Lys Trp
 1           5           10           15

Glu Lys Ile Arg Leu Arg Pro Gly Gly Lys Lys His Tyr Met Leu Lys
          20           25           30

His Leu Val Trp Ala Ser Arg Glu Leu Glu Gly Phe Ala Leu Asn Pro
      35           40           45

Gly Leu Leu Glu Thr Ala Glu Gly Cys Lys Gln Ile Met Lys Gln Leu
 50           55           60

Gln Pro Ala Leu Gln Thr Gly Thr Glu Glu Leu Arg Ser Leu Tyr Asn
 65           70           75           80

Thr Val Ala Thr Leu Tyr Cys Val His Ala Gly Ile Glu Val Arg Asp
          85           90           95

Thr Lys Glu Ala Leu Asp Lys Ile Glu Glu Glu Gln Asn Lys Ser Gln
      100           105           110

Gln Lys Thr Gln Gln Ala Lys Glu Ala Asp Gly Lys Val Ser Gln Asn
 115           120           125

Tyr Pro Ile Val Gln Asn Leu Gln Gly Gln Met Val His Gln Ala Ile
 130           135           140

Ser Pro Arg Thr Leu Asn Ala Trp Val Lys Val Ile Glu Glu Lys Ala
 145           150           155           160

Phe Ser Pro Glu Val Ile Pro Met Phe Thr Ala Leu Ser Glu Gly Ala
      165           170           175

Thr Pro Gln Asp Leu Asn Thr Met Leu Asn Thr Val Gly Gly His Gln
      180           185           190

Ala Ala Met Gln Met Leu Lys Asp Thr Ile Asn Glu Glu Ala Ala Glu
 195           200           205

Trp Asp Arg Leu His Pro Val Gln Ala Gly Pro Val Ala Pro Gly Gln
 210           215           220

Met Arg Asp Pro Arg Gly Ser Asp Ile Ala Gly Ala Thr Ser Thr Leu
 225           230           235           240

Gln Glu Gln Ile Ala Trp Met Thr Ser Asn Pro Pro Val Pro Val Gly
      245           250           255

Asp Ile Tyr Lys Arg Trp Ile Ile Leu Gly Leu Asn Lys Ile Val Arg
      260           265           270

Met Tyr Ser Pro Val Ser Ile Leu Asp Ile Arg Gln Gly Pro Lys Glu
      275           280           285

```

Pro Phe Arg Asp Tyr Val Asp Arg Phe Phe Lys Thr Leu Arg Ala Glu
 290 295 300
 Gln Ala Thr Gln Asp Val Lys Asn Trp Met Thr Glu Thr Leu Leu Val
 305 310 315 320
 Gln Asn Ala Asn Pro Asp Cys Lys Thr Ile Leu Arg Ala Leu Gly Pro
 325 330 335
 Gly Ala Thr Leu Glu Glu Met Met Thr Ala Cys Gln Gly Val Gly Gly
 340 345 350
 Pro Gly His Lys Ala Arg Val Leu Ala Glu Ala Met Ser Gln Ala Asn
 355 360 365
 Ser Val Asn Ile Met Met Gln Lys Ser Asn Phe Lys Gly Pro Arg Arg
 370 375 380
 Asn Val Lys Cys Phe Asn Cys Gly Lys Glu Gly His Ile Ala Lys Asn
 385 390 395 400
 Cys Arg Ala Pro Arg Lys Lys Gly Cys Trp Lys Cys Gly Lys Glu Gly
 405 410 415
 His Gln Met Lys Asp Cys Thr Glu Arg Gln Ala Asn Phe Leu Gly Lys
 420 425 430
 Ile Trp Pro Ser His Lys Gly Arg Pro Gly Asn Phe Leu Gln Asn Arg
 435 440 445
 Ser Glu Pro Ala Ala Pro Thr Val Pro Thr Ala Pro Pro Ala Glu Ser
 450 455 460
 Phe Arg Phe Glu Glu Thr Thr Pro Ala Pro Lys Gln Glu Pro Lys Asp
 465 470 475 480
 Arg Glu Pro Tyr Arg Glu Pro Leu Thr Ala Leu Arg Ser Leu Phe Gly
 485 490 495
 Ser Gly Pro Leu Ser Gln
 500

<210> 23

<211> 849

<212> PRT

<213> Human immunodeficiency virus

<400> 23

Met Arg Val Met Gly Ile Leu Lys Asn Tyr Gln Gln Trp Trp Met Trp
 1 5 10 15
 Gly Ile Leu Gly Phe Trp Met Leu Ile Ile Ser Ser Val Val Gly Asn
 20 25 30
 Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala Lys
 35 40 45

Thr Thr Leu Phe Cys Thr Ser Asp Ala Lys Ala Tyr Glu Thr Glu Val
 50 55 60
 His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
 65 70 75 80
 Gln Glu Ile Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp Lys
 85 90 95
 Asn Asp Met Val Asp Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp
 100 105 110
 Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
 115 120 125
 Lys Cys Arg Asn Val Asn Ala Thr Asn Asn Ile Asn Ser Met Ile Asp
 130 135 140
 Asn Ser Asn Lys Gly Glu Met Lys Asn Cys Ser Phe Asn Val Thr Thr
 145 150 155 160
 Glu Leu Arg Asp Arg Lys Gln Glu Val His Ala Leu Phe Tyr Arg Leu
 165 170 175
 Asp Val Val Pro Leu Gln Gly Asn Asn Ser Asn Glu Tyr Arg Leu Ile
 180 185 190
 Asn Cys Asn Thr Ser Ala Ile Thr Gln Ala Cys Pro Lys Val Ser Phe
 195 200 205
 Asp Pro Ile Pro Ile His Tyr Cys Thr Pro Ala Gly Tyr Ala Ile Leu
 210 215 220
 Lys Cys Asn Asn Gln Thr Phe Asn Gly Thr Gly Pro Cys Asn Asn Val
 225 230 235 240
 Ser Ser Val Gln Cys Ala His Gly Ile Lys Pro Val Val Ser Thr Gln
 245 250 255
 Leu Leu Leu Asn Gly Ser Leu Ala Lys Gly Glu Ile Ile Ile Arg Ser
 260 265 270
 Glu Asn Leu Ala Asn Asn Ala Lys Ile Ile Ile Val Gln Leu Asn Lys
 275 280 285
 Pro Val Lys Ile Val Cys Val Arg Pro Asn Asn Asn Thr Arg Lys Ser
 290 295 300
 Val Arg Ile Gly Pro Gly Gln Thr Phe Tyr Ala Thr Gly Glu Ile Ile
 305 310 315 320
 Gly Asp Ile Arg Gln Ala Tyr Cys Ile Ile Asn Lys Thr Glu Trp Asn
 325 330 335
 Ser Thr Leu Gln Gly Val Ser Lys Lys Leu Glu Glu His Phe Ser Lys
 340 345 350
 Lys Ala Ile Lys Phe Glu Pro Ser Ser Gly Gly Asp Leu Glu Ile Thr
 355 360 365

Thr His Ser Phe Asn Cys Arg Gly Glu Phe Phe Tyr Cys Asp Thr Ser
 370 375 380
 Gln Leu Phe Asn Ser Thr Tyr Ser Pro Ser Phe Asn Gly Thr Glu Asn
 385 390 395 400
 Lys Leu Asn Gly Thr Ile Thr Ile Thr Cys Arg Ile Lys Gln Ile Ile
 405 410 415
 Asn Met Trp Gln Lys Val Gly Arg Ala Met Tyr Ala Pro Pro Ile Ala
 420 425 430
 Gly Asn Leu Thr Cys Glu Ser Asn Ile Thr Gly Leu Leu Leu Thr Arg
 435 440 445
 Asp Gly Gly Lys Thr Gly Pro Asn Asp Thr Glu Ile Phe Arg Pro Gly
 450 455 460
 Gly Gly Asp Met Arg Asp Asn Trp Arg Asn Glu Leu Tyr Lys Tyr Lys
 465 470 475 480
 Val Val Glu Ile Lys Pro Leu Gly Val Ala Pro Thr Glu Ala Lys Arg
 485 490 495
 Arg Val Val Glu Arg Glu Lys Arg Ala Val Gly Ile Gly Ala Val Phe
 500 505 510
 Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met Gly Ala Ala Ser Ile
 515 520 525
 Thr Leu Thr Val Gln Ala Arg Leu Leu Leu Ser Gly Ile Val Gln Gln
 530 535 540
 Gln Asn Asn Leu Leu Arg Ala Ile Glu Ala Gln Gln His Leu Leu Gln
 545 550 555 560
 Leu Thr Val Trp Gly Ile Lys Gln Leu Gln Thr Arg Ile Leu Ala Val
 565 570 575
 Glu Arg Tyr Leu Lys Asp Gln Gln Leu Leu Gly Ile Trp Gly Cys Ser
 580 585 590
 Gly Lys Leu Ile Cys Thr Thr Ala Val Pro Trp Asn Ser Ser Trp Ser
 595 600 605
 Asn Arg Ser His Asp Glu Ile Trp Asp Asn Met Thr Trp Met Gln Trp
 610 615 620
 Asp Arg Glu Ile Asn Asn Tyr Thr Asp Thr Ile Tyr Arg Leu Leu Glu
 625 630 635 640
 Glu Ser Gln Asn Gln Gln Glu Lys Asn Glu Lys Asp Leu Leu Ala Leu
 645 650 655
 Asp Ser Trp Gln Asn Leu Trp Asn Trp Phe Ser Ile Thr Asn Trp Leu
 660 665 670
 Trp Tyr Ile Lys Ile Phe Ile Met Ile Val Gly Gly Leu Ile Gly Leu
 675 680 685

Arg Ile Ile Phe Ala Val Leu Ser Ile Val Asn Arg Val Arg Gln Gly
 690 695 700
 Tyr Ser Pro Leu Pro Phe Gln Thr Leu Thr Pro Asn Pro Arg Glu Pro
 705 710 715 720
 Asp Arg Leu Gly Arg Ile Glu Glu Glu Gly Gly Glu Gln Asp Arg Gly
 725 730 735
 Arg Ser Ile Arg Leu Val Ser Gly Phe Leu Ala Leu Ala Trp Asp Asp
 740 745 750
 Leu Arg Ser Leu Cys Leu Phe Ser Tyr His Arg Leu Arg Asp Phe Ile
 755 760 765
 Leu Ile Ala Ala Arg Val Leu Glu Leu Leu Gly Gln Arg Gly Trp Glu
 770 775 780
 Ala Leu Lys Tyr Leu Gly Ser Leu Val Gln Tyr Trp Gly Leu Glu Leu
 785 790 795 800
 Lys Lys Ser Ala Ile Ser Leu Leu Asp Thr Ile Ala Ile Ala Val Ala
 805 810 815
 Glu Gly Thr Asp Arg Ile Ile Glu Phe Ile Gln Arg Ile Cys Arg Ala
 820 825 830
 Ile Arg Asn Ile Pro Arg Arg Ile Arg Gln Gly Phe Glu Ala Ala Leu
 835 840 845

Gln

<210> 24

<211> 855

<212> PRT

<213> Human immunodeficiency virus

<400> 24

Met Arg Val Arg Gly Ile Leu Arg Ser Trp Gln Gln Trp Trp Ile Trp
 1 5 10 15
 Gly Ile Leu Gly Phe Trp Ile Cys Ser Gly Leu Gly Asn Leu Trp Val
 20 25 30
 Thr Val Tyr Asp Gly Val Pro Val Trp Arg Glu Ala Ser Thr Thr Leu
 35 40 45
 Phe Cys Ala Ser Asp Ala Lys Ala Tyr Glu Lys Glu Val His Asn Val
 50 55 60
 Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro Gln Glu Ile
 65 70 75 80
 Glu Leu Asp Asn Val Thr Glu Asn Phe Asn Met Trp Lys Asn Asp Met
 85 90 95

Val Asp Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp Gln Ser Leu
 100 105 110
 Lys Pro Arg Val Lys Leu Thr Pro Leu Cys Val Thr Leu Lys Cys Thr
 115 120 125
 Asn Tyr Ser Thr Asn Tyr Ser Asn Thr Met Asn Ala Thr Ser Tyr Asn
 130 135 140
 Asn Asn Thr Thr Glu Glu Ile Lys Asn Cys Thr Phe Asn Met Thr Thr
 145 150 155 160
 Glu Leu Arg Asp Lys Lys Gln Gln Val Tyr Ala Leu Phe Tyr Lys Leu
 165 170 175
 Asp Ile Val Pro Leu Asn Ser Asn Ser Ser Glu Tyr Arg Leu Ile Asn
 180 185 190
 Cys Asn Thr Ser Ala Ile Thr Gln Ala Cys Pro Lys Val Ser Phe Asp
 195 200 205
 Pro Ile Pro Ile His Tyr Cys Ala Pro Ala Gly Tyr Ala Ile Leu Lys
 210 215 220
 Cys Lys Asn Asn Thr Ser Asn Gly Thr Gly Pro Cys Gln Asn Val Ser
 225 230 235 240
 Thr Val Gln Cys Thr His Gly Ile Lys Pro Val Val Ser Thr Pro Leu
 245 250 255
 Leu Leu Asn Gly Ser Leu Ala Glu Gly Gly Glu Ile Ile Ile Arg Ser
 260 265 270
 Lys Asn Leu Ser Asn Asn Ala Tyr Thr Ile Ile Val His Leu Asn Asp
 275 280 285
 Ser Val Glu Ile Val Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Gly
 290 295 300
 Ile Arg Ile Gly Pro Gly Gln Thr Phe Tyr Ala Thr Glu Asn Ile Ile
 305 310 315 320
 Gly Asp Ile Arg Gln Ala His Cys Asn Ile Ser Ala Gly Glu Trp Asn
 325 330 335
 Lys Ala Val Gln Arg Val Ser Ala Lys Leu Arg Glu His Phe Pro Asn
 340 345 350
 Lys Thr Ile Glu Phe Gln Pro Ser Ser Gly Gly Asp Leu Glu Ile Thr
 355 360 365
 Thr His Ser Phe Asn Cys Arg Gly Glu Phe Phe Tyr Cys Asn Thr Ser
 370 375 380
 Lys Leu Phe Asn Ser Ser Tyr Asn Gly Thr Ser Tyr Arg Gly Thr Glu
 385 390 395 400
 Ser Asn Ser Ser Ile Ile Thr Leu Pro Cys Arg Ile Lys Gln Ile Ile
 405 410 415

Asp Met Trp Gln Lys Val Gly Arg Ala Ile Tyr Ala Pro Pro Ile Glu
 420 425 430

Gly Asn Ile Thr Cys Ser Ser Ser Ile Thr Gly Leu Leu Leu Ala Arg
 435 440 445

Asp Gly Gly Leu Asp Asn Ile Thr Thr Glu Ile Phe Arg Pro Gln Gly
 450 455 460

Gly Asp Met Lys Asp Asn Trp Arg Asn Glu Leu Tyr Lys Tyr Lys Val
 465 470 475 480

Val Glu Ile Lys Pro Leu Gly Val Ala Pro Thr Glu Ala Lys Arg Arg
 485 490 495

Val Val Glu Arg Glu Lys Arg Ala Val Gly Ile Gly Ala Val Ile Phe
 500 505 510

Gly Phe Leu Gly Ala Ala Gly Ser Asn Met Gly Ala Ala Ser Ile Thr
 515 520 525

Leu Thr Ala Gln Ala Arg Gln Leu Leu Ser Gly Ile Val Gln Gln Gln
 530 535 540

Ser Asn Leu Leu Arg Ala Ile Glu Ala Gln Gln His Met Leu Gln Leu
 545 550 555 560

Thr Val Trp Gly Ile Lys Gln Leu Gln Ala Arg Val Leu Ala Ile Glu
 565 570 575

Arg Tyr Leu Lys Asp Gln Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly
 580 585 590

Lys Leu Ile Cys Thr Thr Thr Val Pro Trp Asn Ser Ser Trp Ser Asn
 595 600 605

Lys Thr Gln Gly Glu Ile Trp Glu Asn Met Thr Trp Met Gln Trp Asp
 610 615 620

Lys Glu Ile Ser Asn Tyr Thr Gly Ile Ile Tyr Arg Leu Leu Glu Glu
 625 630 635 640

Ser Gln Asn Gln Gln Glu Gln Asn Glu Lys Asp Leu Leu Ala Leu Asp
 645 650 655

Ser Arg Asn Asn Leu Trp Ser Trp Phe Asn Ile Ser Asn Trp Leu Trp
 660 665 670

Tyr Ile Lys Ile Phe Ile Met Ile Val Gly Gly Leu Ile Gly Leu Arg
 675 680 685

Ile Ile Phe Ala Val Leu Ser Ile Val Asn Arg Val Arg Gln Gly Tyr
 690 695 700

Ser Pro Leu Ser Phe Gln Thr Leu Thr Pro Asn Pro Arg Gly Leu Asp
 705 710 715 720

Arg Leu Gly Arg Ile Glu Glu Glu Gly Gly Glu Gln Asp Arg Asp Arg
 725 730 735

Ser Ile Arg Leu Val Gln Gly Phe Leu Ala Leu Ala Trp Asp Asp Leu
740 745 750

Arg Ser Leu Cys Leu Phe Ser Tyr His Arg Leu Arg Asp Leu Ile Leu
755 760 765

Val Thr Ala Arg Val Val Glu Leu Leu Gly Arg Ser Ser Pro Arg Gly
770 775 780

Leu Gln Arg Gly Trp Glu Ala Leu Lys Tyr Leu Gly Ser Leu Val Gln
785 790 795 800

Tyr Trp Gly Leu Glu Leu Lys Lys Ser Ala Thr Ser Leu Leu Asp Ser
805 810 815

Ile Ala Ile Ala Val Ala Glu Gly Thr Asp Arg Ile Ile Glu Val Ile
820 825 830

Gln Arg Ile Tyr Arg Ala Phe Cys Asn Ile Pro Arg Arg Val Arg Gln
835 840 845

Gly Phe Glu Ala Ala Leu Gln
850 855

<210> 25
<211> 20
<212> PRT
<213> Human immunodeficiency virus

<400> 25
Asp Ile Lys Gln Gly Pro Lys Glu Pro Phe Arg Asp Tyr Val Asp Arg
1 5 10 15

Phe Phe Lys Thr
20

<210> 26
<211> 60
<212> DNA
<213> Human immunodeficiency virus

<400> 26
gacataaaac aaggaccaaa agagcccttt agagactatg tagaccggtt ctttaaaacc 60

<210> 27
<211> 20
<212> PRT
<213> Human immunodeficiency virus

<400> 27
Asp Ile Arg Gln Gly Pro Lys Glu Pro Phe Arg Asp Tyr Val Asp Arg
1 5 10 15

Phe Phe Lys Thr
20

<210> 28
<211> 47
<212> PRT
<213> Human immunodeficiency virus

<400> 28
Thr Ile Thr Ile Thr Cys Arg Ile Lys Gln Ile Ile Asn Met Trp Gln
1 5 10 15
Lys Val Gly Arg Ala Met Tyr Ala Pro Pro Ile Ala Gly Asn Leu Thr
20 25 30
Cys Glu Ser Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly
35 40 45

<210> 29
<211> 48
<212> PRT
<213> Human immunodeficiency virus

<400> 29
Ser Ile Ile Thr Leu Pro Cys Arg Ile Lys Gln Ile Ile Asp Met Trp
1 5 10 15
Gln Lys Val Gly Arg Ala Ile Tyr Ala Pro Pro Ile Glu Gly Asn Ile
20 25 30
Thr Cys Ser Ser Ser Ile Thr Gly Leu Leu Leu Ala Arg Asp Gly Gly
35 40 45